

STIC-Biotech/Ch mLib

83753

From: Chan, Christina  
Sent: Tuesday, January 07, 2003 10:25 AM  
To: Duffy, Patricia; STIC-Biotech/ChemLib  
Subject: RE: PLEASE RUSH SEQUENCE SEARCH--- Amendment due this Biweek  
Importance: High

Please rush: Thanks Chris

Chris Chan  
TC 1600 New Hire Training Coordinator and SPE 1644  
308-3973  
CM-1, 9B19

-----Original Message-----

From: Duffy, Patricia  
Sent: Monday, January 06, 2003 10:43 AM  
To: Chan, Christina  
Subject: PLEASE RUSH SEQUENCE SEARCH--- Amendment due this Biweek  
Importance: High

In re: 09/438,185

Please search SEQ ID NO:1047.  
Please search na residues 1200675-1199590 of SEQ ID NO:1.  
Please include an interference search.  
Please print out top 50 hits in each category.

Thank you.

~~Patricia A. Duffy~~  
~~CM1-8B05~~

AU 1645  
703-305-7555

POINT OF CONTACT:  
PAUL SCHULWITZ  
TECHNICAL INFO. SPECIALIST  
CM1 6B06 TEL. (703) 305-1954

Searcher: \_\_\_\_\_  
Name: \_\_\_\_\_  
Position: \_\_\_\_\_  
Date Picked Up: 1/7  
Date Completed: 1/10  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 11:41:02 ; Search time 45 Seconds  
(without alignments)  
10621.161 Million cell updates/sec

Title: US-09-438-185A-1\_COPY\_1199590\_1200675

Perfect score: 1086  
Sequence: 1 ttggcaagtagctcaaac.....gtctctcgcagtagtgac 1086

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 389086 seqs, 220051671 residues

Total number of hits satisfying chosen parameters: 778172

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Published\_Applications\_NA:\*

1: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Description	
Result No.	Score	Query Match	Length DB ID		
C 1	46.6	4.3	1642	10	US-09-205-658-312
C 2	41.4	3.8	817	10	US-09-205-658-314
C 3	40.6	3.7	640681	10	US-09-790-988-5
C 4	36.6	3.4	2098	10	US-09-971-388-5
C 5	36.4	3.4	1241	10	US-09-737-626A-25
C 6	36.4	3.4	2000	9	US-09-938-842A-3348
C 7	36.4	3.4	2004	9	US-09-887-576-256
C 8	36.4	3.4	326014	10	US-09-731-231A-3
C 9	36.2	3.3	580	10	US-09-920-300A-111
C 10	36.2	3.3	1985	10	US-09-817-913-4
C 11	36.2	3.3	33	10	US-09-817-913-4
C 12	36.2	3.3	18998	10	US-09-764-869-1952
C 13	36.2	3.3	32191	10	US-09-764-869-1955
C 14	36.2	3.3	1128	10	US-09-815-242-4614
C 15	36	3.3	1134	10	US-09-815-242-8602
C 16	35.8	3.3	7311	12	US-10-044-090-455
C 17	35.8	3.3	952	10	US-09-764-877-2599
C 18	35.4	3.3	952	10	US-09-764-877-2599
C 19	35.4	3.3	1134	10	US-09-815-242-8628

C 20	35.2	3.2	271	10	US-09-777-564-1653
C 21	34.8	3.2	335913	9	US-09-754-853A-2
C 22	34.8	3.2	335913	9	US-09-754-853A-3
C 23	34.6	3.2	1391	10	US-09-925-301-66
C 24	34.6	3.2	3092	10	US-09-925-302-310
C 25	34.4	3.2	10195	10	US-09-764-864-1600
C 26	34.2	3.1	4201	10	US-09-925-301-478
C 27	34	3.1	35641	10	US-09-962-436-306
C 28	34	3.1	35641	10	US-09-880-107-2225
C 29	33.8	3.1	180557	12	US-10-003-806-6
C 30	33.8	3.1	180557	12	US-10-003-806-9
C 31	33.6	3.1	147309	10	US-09-743-312-3
C 32	33.4	3.1	2679	10	US-09-880-107-2219
C 33	33.4	3.1	2680	10	US-09-782-378A-20
C 34	33.4	3.1	76056	9	US-10-109-551-1
C 35	33.2	3.1	568	10	US-09-964-761-0781
C 36	33.2	3.1	1738	10	US-09-764-864-1784
C 37	33.2	3.1	2000	9	US-09-938-842A-4741
C 38	33.2	3.1	198285	10	US-09-880-107-3814
C 39	32.8	3.0	328	10	US-09-878-574-3709
C 40	32.8	3.0	406	10	US-09-878-574-3709
C 41	32.8	3.0	548	10	US-09-924-035A-596
C 42	32.8	3.0	806	10	US-09-939-980-6
C 43	32.8	3.0	31412	9	US-10-109-551-3
C 44	32.6	3.0	601	9	US-09-854-133-294
C 45	32.6	3.0	601	10	US-09-738-973-294
C 46	32.6	3.0	2000	9	US-09-938-842A-3797
C 47	32.4	3.0	303	10	US-09-969-708-291
C 48	32.4	3.0	1658	10	US-09-925-302-228
C 49	32.4	3.0	4436	9	US-09-522-334-26
C 50	32.4	3.0	513509	9	US-09-754-853A-4
C 51	32.2	3.0	2000	9	US-09-938-842A-4311
C 52	32.2	3.0	23419	10	US-09-764-869-1364
C 53	32.2	3.0	26006	10	US-09-764-869-1963
C 54	32.2	3.0	26006	10	US-09-764-864-1638
C 55	32.2	3.0	26013	10	US-09-764-869-1961
C 56	32.2	3.0	26013	10	US-09-764-864-1636
C 57	32.2	3.0	26018	10	US-09-764-869-1962
C 58	32.2	3.0	26018	10	US-09-764-864-1637
C 59	32.2	3.0	302350	10	US-09-962-832-154
C 60	32	2.9	470	10	US-09-867-701-78
C 61	32	2.9	464	10	US-09-969-708-534
C 62	32	2.9	570	9	US-08-788-827-84
C 63	32	2.9	580	9	US-08-788-827-79
C 64	32	2.9	768	9	US-10-205-193-162
C 65	32	2.9	828	10	US-09-925-300-315
C 66	32	2.9	954	10	US-09-947-971-3
C 67	32	2.9	1079	10	US-09-925-300-164
C 68	32	2.9	1152	9	US-10-012-896-997
C 69	32	2.9	1173	10	US-09-880-107-2297
C 70	32	2.9	1595	9	US-10-051-307-1
C 71	32	2.9	2000	9	US-09-938-842A-4033
C 72	32	2.9	2000	9	US-09-938-842A-4200
C 73	32	2.9	4857	10	US-09-764-847-1929
C 74	32	2.9	8220	9	US-10-153-273-11
C 75	32	2.9	368004	10	US-09-949-654-3
C 76	31.8	2.9	1253	9	US-10-002-344A-99
C 77	31.8	2.9	32249	10	US-09-764-869-2314
C 78	31.6	2.9	330	10	US-09-983-965-2145
C 79	31.6	2.9	375	9	US-09-796-692-4292
C 80	31.6	2.9	396	9	US-09-970-966-41
C 81	31.6	2.9	396	10	US-09-825-294-41
C 82	31.6	2.9	492	10	US-09-954-456-46
C 83	31.6	2.9	492	10	US-09-954-456-962
C 84	31.6	2.9	495	10	US-09-783-590-2867
C 85	31.6	2.9	519	10	US-09-864-761-6821
C 86	31.6	2.9	1643	9	US-10-001-887-41
C 87	31.6	2.9	1767	9	US-09-989-919-21
C 88	31.6	2.9	2000	9	US-09-832-849A-349
C 89	31.6	2.9	2000	9	US-09-938-842A-165
C 90	31.6	2.9	6823	10	US-09-954-456-1215
C 91	31.6	2.9	13957	10	US-09-782-378A-22
C 92	31.6	2.9	13957	10	US-09-880-107-2284

Sequence 1653, Ap  
Sequence 2, Appli  
Sequence 3, Appli  
Sequence 66, Appli  
Sequence 310, App  
Sequence 1600, App  
Sequence 478, App  
Sequence 306, App  
Sequence 2225, App  
Sequence 6, Appli  
Sequence 3, Appli  
Sequence 2219, Ap  
Sequence 20, Appli  
Sequence 1, Appli  
Sequence 9781, Ap  
Sequence 1784, Ap  
Sequence 4741, Ap  
Sequence 3814, Ap  
Sequence 733, App  
Sequence 3709, Ap  
Sequence 396, App  
Sequence 6, Appli  
Sequence 3, Appli  
Sequence 294, App  
Sequence 294, App  
Sequence 3797, Ap  
Sequence 291, App  
Sequence 228, App  
Sequence 26, Appli  
Sequence 4, Appli  
Sequence 4311, Ap  
Sequence 1364, Ap  
Sequence 1963, Ap  
Sequence 1638, Ap  
Sequence 1961, Ap  
Sequence 1636, Ap  
Sequence 1962, Ap  
Sequence 1637, Ap  
Sequence 154, App  
Sequence 78, Appli  
Sequence 534, App  
Sequence 79, Appli  
Sequence 162, App  
Sequence 315, App  
Sequence 3, Appli  
Sequence 164, App  
Sequence 997, App  
Sequence 2297, Ap  
Sequence 10, Appli  
Sequence 4033, Ap  
Sequence 4200, Ap  
Sequence 1529, Ap  
Sequence 11, Appli  
Sequence 3, Appli  
Sequence 99, Appli  
Sequence 2314, Ap  
Sequence 2145, Ap  
Sequence 4292, Ap  
Sequence 41, Appli  
Sequence 41, Appli  
Sequence 46, Appli  
Sequence 962, App  
Sequence 2867, Ap  
Sequence 6821, Ap  
Sequence 21, Appli  
Sequence 349, App  
Sequence 4165, App  
Sequence 1215, Ap  
Sequence 22, Appli  
Sequence 2284, Ap





us-09-438-185a-1\_1199590\_1200675.rnpb

Fri Jan 10 12:01:28 2003

```

; EARLIER APPLICATION NUMBER: 08/857,076
; EARLIER FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: 08/888,534
; EARLIER FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: 08/908,000
; EARLIER FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 328
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 314
; LENGTH: 817
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
; US-09-205-658-314

Query Match      3.8%; Score 41.4; DB 10; Length 817;
Best Local Similarity 45.2%; Pred. No. 0.28;
Matches 239; Conservative 0; Mismatches 281; Indels 9; Gaps 2;

QY 113 TCATCAAAATGCTTATTGAAATAAAGTCTCTGTGAGTGTGATGTAAGGGAAGA 172
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 639 TCTTCAAAATTTCTAGTATAGTAATACGCTGACTGGAAGTAGTATAGACATCTTGC 580
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 173 CGAATAATCTGATCCCAATCTTAAGGGAAGACACAGTACGTTATCAATGAAGCGTGCA 232
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 579 TCACACACAGATCCGATCAAAACGAATAATGTTGCACCTACCCCTCAAGCGCATTTGC 520
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 233 AGTTCTGAGAGACATCAATGAAGAGCGCTCCATATGCTTTTCTTCTTCTTCTTCTTCA 292
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 519 AACTGCCAGACATGCTCAGAGTCTGCTCGTAFCTTTAAATCTTTCATGATTGAT 460
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 293 ATAAGTCCGCTTTCACAGTAACCAAGACAGAGTACAAATAGCGATCAGATTGCT 352
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 459 CCATTTCTTTTACTGGAG-AATCGGCGGCTCATCAGCAGACAGACAAATCAATGA 401
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 353 AGGTTGATGCGTGTGTTTACTAGGAGAGCTGTGATCTTTTCTTCTTCTTCTTCTTCTG 412
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 400 AAAGAAGTAGAGTGTCAGAGCTTCTTCAAA-----TCTTCTCTGATGCTCCAGA 349
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 413 AAGAGCTTCCCATGTTTATGAAATATCAAGAAATGAGGAGATGAGAACCAAGCAGC 472
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 348 GAAGCTAATCCAATCTCTTGAGAAATAGCAAAATCTGGAATCAGCAATAGAGCCATG 289
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 473 TGCCTTAAAGCTGATGATGAGATCAGGAGTTAAGGAGAAATTTATCTTTATCGAGATT 532
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 288 TGACCCATGAGCTGTAACGCTGTCTGTTCTGAGTGTAAATGATGCGCATGATGG 229
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 533 CGCATTACAGAGCAATGGGAAATGACGCTCTGCAACAGACAGCAAGTATTGATGGGA 592
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 228 CGRACGTATTGATGTCAGAAGACACAGATATGCAAGACCTGCCAAGAAATCAGAGCT 169
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 593 GCTACAAATCCTGAACGGGATAATAGGAAATATGTTTCTTAATCGA 641
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 168 GATAAGTATCCGGAGCTGGGCGACACAGCAATCCAGTTTGTGACACTA 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: MATSUDA, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; PRIOR FILING DATE: 2001-02-23
; CURRENT APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1

```

```

; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
; US-09-790-988-1

Query Match      3.7%; Score 40.6; DB 10; Length 640681;
Best Local Similarity 43.4%; Pred. No. 7.3;
Matches 187; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

QY 119 AAATGCTTATTGAAATAAAGTCTCTTGTGGAGTTGATGTATTGAGGGAAGAGCAATA 178
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190122 AATAATGATTTTATAAAAAAGTTGATAGTTTAAAAAAAATAAAAAATGAAAAACA 190181
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 179 ATCTGATCCAATTTAAAGGGAGAACACGCTACGTATCAATGAAAGCGTCCCAAGTTCC 238
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190182 ATAGTGCCTAAAAAATATTAAAAAATATTTTCTCAATTATAGCCTGATTCCTTGG 190241
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 239 TGAGGAGAACTGATGAAGACGCTCCATATGCTTTCTTCTTCTTCTTCTTCTTCAAT 298
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190242 AATCGAAACATCGTACACACACGCTGTAATAGATCTTTTAAATTTTAAATATACA 190301
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 299 CCGTTTCAACAGTAACCAAGACGCTACATAGCGATCAGATTCCTTTGTAGGTT 358
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190302 ACTAATTTAGTACAAAAAATCATATCATAGAGAAATAATCATATTCGATACGCAT 190361
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 359 TGTATGCTGTTTCTTACTAGGAAGAGCTTGTACTTTTCTTATGACTTTAGTAGAGT 418
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190362 CATCTAAATTTTAAATAAACACCTTTTAAATACTTTAGTAATAGTTTCTTTTCT 190421
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 419 CTTCCTGATTTTATGAAAAATTCAGAAAAATGAGGAGTGTAGAACCAAGCGACGTCCT 478
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190422 CTGATGCTTTTAAATGTTTAAATGTAATGTAATGTAATGTAATGTAATGTAAT 190481
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 479 AAAAGGTCATGGATGAGATCAGGAGTTAAGGAGAAATTTATCTTTATCGAGAGTTCCG 538
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190482 ATATTAAAGGGAATTTAAATAAAAAAATAATATGCTTCTTTATATAAATTTTAA 190541
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 539 ACAGAGGCAAT 549
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190542 AAGAATCTAT 190552
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
US-09-871-388-5/C
; Sequence 5, Application US/09871388
; Patent No. US20020127621A1
; GENERAL INFORMATION:
; APPLICANT: Rubin, Gerald M.
; APPLICANT: Pan, Duojia
; APPLICANT: Rooke, Jenny
; APPLICANT: Yavari, Reza
; APPLICANT: Xu, Tian
; TITLE OF INVENTION: KUZ: A No. US20020127621A1 Family of Metalloproteases
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/871,388
; FILING DATE: 31-May-2001
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/937,931
; FILING DATE: <unknown>

```

```
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B97-081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2098 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-871-388-5

Query Match          3.4%; Score 36.6; DB 10; Length 2098;
Best Local Similarity 65.1%; Pred. No. 7;
Matches 54; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 816 AGCAACCTCCCTTTGATTCCTTTGATATGACAGCAAGCAACAAAGACTAAGTTTGA 875
      ||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1136 AGCCACAAAAGCTTTGAAGGTTTGATATTCAGAGAAGATGACCAAGAGCTTCTTTATT 1077
      ||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 876 AAAATCTTAATAGAAATGCAT 898
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1076 AGGCTACTGAATATTAATTTAT 1054
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
US-09-737-626A-25
; Sequence 25, Application US/09737626A
; Patent No. US2002014304A1
; GENERAL INFORMATION:
; APPLICANT: Fincher, Karen
; APPLICANT: Flasinski, Jack
; APPLICANT: Wilkinson, Jack
; TITLE OF INVENTION: No. US2002014304A1 Plant Expression Constructs
; FILE REFERENCE: 38-21(51499)B
; CURRENT APPLICATION NUMBER: US/09/737,626A
; CURRENT FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: 09/737,626
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent version 3.0
; SEQ ID NO 25
; LENGTH: 1241
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1241)
; OTHER INFORMATION: n= a or g or c or t/u
; OTHER INFORMATION: Act7 promoter polynucleotide sequence and intron
US-09-737-626A-25

Query Match          3.4%; Score 36.4; DB 10; Length 1241;
Best Local Similarity 59.8%; Pred. No. 6.3;
Matches 61; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 361 TATCGGTGTTGTTTACTAGGAAGCTGTGACTTTTCTATGACTTTAGTGAAGTCT 420
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1030 TAGGTGTTGATGTTTACAAGTGAATCTAGTGTCTCTCTTGAGATCTGTGAAGTTGA 1089
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 421 TCCCATGTTTATGAAAATTCAGAAATGAGGATGTAGAAG 462
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1090 ACCTAGTTTCTCAATAATCAACATATGAAGCGATGTTTGAG 1131
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
US-09-938-842A-3348
; Sequence 3348, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRI1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3348
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3348

Query Match          3.4%; Score 36.4; DB 9; Length 2000;
Best Local Similarity 59.8%; Pred. No. 7.8;
Matches 61; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 361 TATCGGTGTTTACTAGGAAGCTGTGACTTTTCTATGACTTTAGTGAAGTCT 420
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1789 TAGGTGTTGATGTTTACAAGTGAATCTAGTGTCTCTCTTGAGATCTGTGAAGTTGA 1848
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 421 TCCCATGTTTATGAAAATTCAGAAATGAGGATGTAGAAG 462
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1849 ACCTAGTTTCTCAATAATCAACATATGAAGCGATGTTTGAG 1890
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
US-09-887-576-256
; Sequence 256, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, P.
; APPLICANT: Brown, D.
; APPLICANT: Chang, H.
; APPLICANT: Zhu, T.
; APPLICANT: Han, B.
; APPLICANT: Wang, X.
; APPLICANT: Cooper, Bret
; TITLE OF INVENTION: Promoters for regulation of plant expression
; FILE REFERENCE: 1360.001US1
; CURRENT APPLICATION NUMBER: US/09/887,576
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/213,848
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,087
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/258,692
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 875
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 256
; LENGTH: 2004
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-887-576-256

Query Match          3.4%; Score 36.4; DB 10; Length 2004;
Best Local Similarity 59.8%; Pred. No. 7.8;
Matches 61; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 361 TATCGGTGTTTACTAGGAAGCTGTGACTTTTCTATGACTTTAGTGAAGTCT 420
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Fri Jan 10 12:01:28 2003

```

Db 1789 TAGGTGTTGATGTTTACAAAGTGAATTCAGTGTCTTCTTTGAGATCGTGAAGTTGA 1848
QY 421 TCCCATGTTTATGAAAATTCAGAAAATGAGGATGAGAG 462
Db 1849 ACCTAGTTTCTCAATAATCAACATATGAAGCATGTTTGAG 1890

RESULT 8
US-09-731-231A-3/c
; Sequence 3, Application US/09731231A
; Patent No. US20020082189A1
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CLO01000
; CURRENT APPLICATION NUMBER: US/09/731,231A
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 326014
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(326014)
; OTHER INFORMATION: n = A,T,C or G
US-09-731-231A-3

Query Match 3.4%; Score 36.4; DB 10; Length 326014;
Best Local Similarity 52.7%; Pred. No. 65;
Matches 79; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 755 AGGAGTCGATACCATAAATTCGATTAAAGGAGTGTCTTACACAGCCATACAGTG 814
Db 42622 AGATGACAATAAAATCAAGATACAGTAAATTTAGAGAGCAACAAAGCTGA 42563
QY 815 GAGCACTTCCCTTTGATTCTTCATATGACAGAGATGACAAAGAGTAAGTTTGG 874
Db 42562 TTAGCTCCATTCCTTGATAGTAAATGAGAAAGATAACACTGTTTAAAGTAAA 42503
QY 875 AAAATCTTAATAGAAATGCAATTTGT 904
Db 42502 TTAATATATCTAGCAAGAGTTGTT 42473

RESULT 9
US-09-920-300A-111
; Sequence 111, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920,300A
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 111
; LENGTH: 580
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-920-300A-111

Query Match 3.3%; Score 36.2; DB 10; Length 580;
Best Local Similarity 53.1%; Pred. No. 5.2;
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```

```

QY 799 AACGCCATATCAGTGAGCAACCTTCCCTTTGATTTTGTATTAATGACAGAGATGAAC 858
Db 383 ATCAGCCACATTTCTCGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 442
QY 859 AAGAGTAAGTTTTTGAAAAATCTAAATAGAAATGCAATTTGTGTCGAGCTAAAGCTTG 918
Db 443 ACAAGCTATCCGCTTGTCTGATGCTCAATAGAAAATCTCTTGTCTGATCTTCTCCATC 502
QY 919 CTTCCTCTTATTTTCTTTTGTAGA 943
Db 503 TTCTATCTCCACTGTCTTTCATGAACA 527

RESULT 10
US-10-033-528-111
; Sequence 111, Application US/10033528
; Patent No. US20020131971A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547C1
; CURRENT APPLICATION NUMBER: US/10/033,528
; CURRENT FILING DATE: 2001-12-26
; NUMBER OF SEQ ID NOS: 1896
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 111
; LENGTH: 580
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-033-528-111

Query Match 3.3%; Score 36.2; DB 12; Length 580;
Best Local Similarity 53.1%; Pred. No. 5.2;
Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 799 AACGCCATATCAGTGAGCAACCTTCCCTTTGATTTTGTATTAATGACAGAGATGAAC 858
Db 383 ATCAGCCACATTTCTCGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 442
QY 859 AAGAGTAAGTTTTTGAAAAATCTAAATAGAAATGCAATTTGTGTCGAGCTAAAGCTTG 918
Db 443 ACAAGCTATCCGCTTGTCTGATGCTCAATAGAAAATCTCTTGTCTGATCTTCTCCATC 502
QY 919 CTTCCTCTTATTTTCTTTTGTAGA 943
Db 503 TTCTATCTCCACTGTCTTTCATGAACA 527

RESULT 11
US-09-817-913-4/c
; Sequence 4, Application US/09817913
; Patent No. US20020061860A1
; GENERAL INFORMATION:
; APPLICANT: Bonfils, Claire
; APPLICANT: Besterman, Jeffrey
; APPLICANT: Inhibition of Specific Histone Deacetylase Isoforms
; TITLE OF INVENTION: 106101.145
; FILE REFERENCE: 106101.145
; CURRENT APPLICATION NUMBER: US/09/817,913
; CURRENT FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: US 60/192,157
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1985
; TYPE: DNA
; ORGANISM: Human

```



us-09-438-185a-1\_1199590\_1200675.rnpb

Fri Jan 10 12:01:28 2003

```

; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 4614
; LENGTH: 1128
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; US-09-815-242-8602

```

```

Query Match 3.3% Score 36; DB 10; Length 1128;
Best Local Similarity 50.0%; Pred. No. 7.7; Indels 0; Gaps 0;
Matches 90; Conservative 0; Mismatches 0;

Qy 35 TCTTGATTGTAAGGGGAATGATCTACAGACACCTTGGTCGAGCATCCATCTTAATTT 94
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 832 TCTTTATTAGCAGCGTAAGGGTGAATACGTTATTAAATTAAGTACAGGTGATGCGTT 891
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 95 GAAGTGAGTTCTACAGTTTCATCAAAATGCTTATTGAAATAAAGTCTCTGTGGAGTT 154
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 892 GTTGAATATCTCTACATACACCAAGAAATGGTACTGTAAAGAGAAAGTGTATGCAAC 951
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 155 GATGATTGAAGGGAAGACGAATAATCTGATCCAAATCTTAAAGGGAGAACACGTACGTTA 214
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 952 GATGTTGAAGGTGGTAGCCTGAAATTTTCATCACTCTAAATATATGATGATGCTTTA 1011
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 16
US-09-815-242-8602
; Sequence 8602, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Frawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8602

```

```

; LENGTH: 1134
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1).....(1134)
; US-09-815-242-8602

Query Match 3.3% Score 36; DB 10; Length 1134;
Best Local Similarity 50.0%; Pred. No. 7.7; Indels 0; Gaps 0;
Matches 90; Conservative 0; Mismatches 0;

Qy 35 TCTTGATTGTAAGGGGAATGATCTACAGACACCTTGGTCGAGCATCCATCTTAATTT 94
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 835 TCTTTATTAGCAGCGTAAGGGTGAATACGTTATTAAATTAAGTACAGGTGATGACGTT 894
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 95 GAAGTGAGTTCTACAGTTTCATCAAAATGCTTATTGAAATAAAGTCTCTGTGGAGTT 154
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 895 GTTGAATATCTCTACATACCAAGAAATGGTACTGTAAAGAGAAAGTGTATGCAAAC 954
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 155 GATGATTGAAGGGAAGACGAATAATCTGATCCAAATCTTAAAGGGAGAACACGTACGTTA 214
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 955 GATGTTGAAGGTGGTAGCCTGAAATTTTCATCACTCTAAATATATGATGATGCTTTA 1014
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 17
US-10-044-090-455
; Sequence 455, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 455
; LENGTH: 7311
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 347926.5
; NAME/KEY: unsure
; LOCATION: 5438, 5455, 5470, 5482, 5495, 5512, 5559
; OTHER INFORMATION: a, t, c, g, or other
; US-10-044-090-455

```

```

Query Match 3.3% Score 35.8; DB 12; Length 7311;
Best Local Similarity 50.9%; Pred. No. 19; Indels 0; Gaps 0;
Matches 85; Conservative 0; Mismatches 82;

Qy 792 TGTTCACACACCCATATCAGTGGAGCAACCTTCCCTTTGATTCTTTGATTAATGACAGAA 851
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2391 TGATTACTGTAGTCATATTTATGAAAAAGGTTTGTGTTTACTCTTCTAGTAGAGAAA 2450
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 852 GATGAACAAAGAGTAAGTTTGTAAAAAATCTTAAATAAGAAATGCAATTTGTGCGAGCTA 911
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2451 GTGGGACAAAATATATCTTTTGAATAAATGCTATATGCGACCTAAATATTTTCTTTT 2510
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 912 AAGCTTGCTTCTCTTTTATTTTCCCTTTTGTAGAATGATTCGGTAGTAG 958
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2511 AAAATGCCTTAAGTTGCAGTCTCATTTTGTATAATCATTTTGTCTCCAG 2557
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 18
US-09-764-877-2599/C
; Sequence 2599, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

```

```
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2599
; LENGTH: 952
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-2599

Query Match      3.3%; Score 35.4; DB 10; Length 952;
Best Local Similarity 47.9%; Pred. No. 10;
Matches 102; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 712 AAATCTGGACAATAGCTTTTCCATAGAGAAACCCGTGAGGAAGAGTGCATACCATAA 771
Db 259 ATATATAGGAAACTCCCTTTAAATAAGTGTAATTCATAAAGATGAAGTGAATAAAGA 200

QY 772 ATTCGGATTAAAAAGGGAGTGCTTTACAAAGCCATATCATGTGGAGCAACCTTCCCTTTG 831
Db 199 TGACAGAAATAGAGCAATCTAATTTTCAATGCAAAATCATATGTGAAAAAATGAACACAT 140

QY 832 ATTCCTTGTAATGACAGAAAGATGAACAAAGAGTAAGTTTTTTGAAAAAATCTAAATAAGA 891
Db 139 AATTTAAATAAATAATAGTAGTAGATAGATGACCATGTCAACTAAATCTCTTAAATAACA 80

QY 892 AATGCATTTTGTGCGAGCTAAAGCTTGTCTCTC 924
Db 79 AAAGTCTCACTCATATATAAACTTATACCTC 47

RESULT 19
US-09-815-242-8628
; Sequence 8628, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8628
; LENGTH: 1134
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS

; LOCATION: (1)...(1134)
US-09-815-242-8628

Query Match      3.3%; Score 35.4; DB 10; Length 1134;
Best Local Similarity 49.7%; Pred. No. 11;
Matches 90; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 34 CTCTTTGATTGTAAGGGGAATTGATTTCAACAGACACCTTGTGTCGAGCATCCATTCTTAATTT 93
Db 834 CTCCTTTATTAGCGCGTGAAGGTGGTAATAACGTTATTAAATTAAGTACAGGTGATGCGT 893

QY 94 TGAAGTCAGTCTTACCAGTTCATCAAAATGCTTTATTGAAAATAAAGTCTCTTGTGGAGT 153
Db 894 TGTGAATTCCTCTTCATCATCACCAGAAATTTGGTACTGTAAAGAAAGATTGATGCCAAA 953

QY 154 TGATGTATTGAAGGGGAAGACGAATAATCTGTATCCCAANTCTAAAGGGGAGAACACGTACGTT 213
Db 954 CGATGTTGAAGGTGGTAGCCTGAAAATTTTCATCAACTCTAAAATATATATGATGATGCTTT 1013

QY 214 A 214
Db 1014 A 1014

RESULT 20
US-09-777-564-1653/c
; Sequence 1653, Application US/09777564
; Patent No. US20020022591A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.493
; CURRENT APPLICATION NUMBER: US/09/777,564
; CURRENT FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 1730
; SOFTWARE: FastSeq for Window Version 4.0
; SEQ ID NO 1653
; LENGTH: 271
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(271)
; OTHER INFORMATION: n = A,T,C or G
US-09-777-564-1653

Query Match      3.2%; Score 35.2; DB 10; Length 271;
Best Local Similarity 52.4%; Pred. No. 6.8;
Matches 76; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 799 AACAGCCATATCAGTGGAGCAACCTTCCCTTTGATTTCTTTGATAATGACAGAGATGAAC 858
Db 248 ATCAGCCACNTTCTTCGACCTCTTCTCTCATCTCAGAAATCTGAGAAATCTTTCATC 189

QY 859 AAGAGCTAAGTTTTTTGAAAAAATCTTAAATAGAAATGCATTTGTGTGAGCTTAAAGCTTG 918
Db 188 ACAAGCTATCCGCTTGTCTGTGATGCTCGAATAGAAATTTCTCTGTGTGATCTTCTCCATC 129

QY 919 CTCTCTCTTATTCTTCTTTTGTAGA 943
Db 128 TTCATCTCCACTGTCTTTCATGAACA 104

RESULT 21
US-09-754-853A-2
; Sequence 2, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
```

APPLICANT: Wang, Ming Li  
TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
FILE REFERENCE: 38-10(15810)B  
CURRENT APPLICATION NUMBER: US/09/754,853A  
CURRENT FILING DATE: 2001-01-05  
PRIOR APPLICATION NUMBER: US 60/174,880  
PRIOR FILING DATE: 2000-01-07  
NUMBER OF SEQ ID NOS: 1119  
SEQ ID NO 2  
LENGTH: 335913  
TYPE: DNA  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (45163)..(45314),(45450)..(45509),(46941)..(48763),(48975)..(49573)  
OTHER INFORMATION: Clone ID: 240017\_region\_G3  
US-09-754-853A-2

Query Match 3.2%; Score 34.8; DB 9; Length 335913;  
Best Local Similarity 44.9%; Pred. No. 1.6e+02;  
Matches 132; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 574 AGACAAGTATTGATGGGAGCTCAAAATCCTGAACGGGATATAGGAAATGCTTTC 633  
DB 199825 AAACGACGATGCTTTGCTATAGGACAGTTGCCAATTGAAGATACAAAGTCTTGACC 199884  
QY 634 TAATTCGAAGAATTAAATGACTGCTTGAAGAAATCAGAAAGGAGCCGAAAGC 693  
DB 199885 TTAACCTTTGCAATTCATCATCTGCGAAATCAACGACTAAAGAAACCGTATGC 199944  
QY 694 TTCTAAGTAATCAAGAAATCTTGGACAATAGCTTTTCCATAGAGAAACCGTGAGGA 753  
DB 199945 ATTTAGAATGTAAATAATTTGGTGTCTATATCACTTAATTAAGGAAAG 200004  
QY 754 AAGGAGTCGATACCAATTAATTCGCAATAGCTTTTCCATAGAGAAACCGTGAGGA 813  
DB 200005 AGTGTACAGGTACAAACAATTTGAACCTTTTCAATTAATTAATGCGATAT 200064  
QY 814 GGAGCAACCTCCCTTTGATCTTTGATATGACAGAGATGACAAAGAGTAA 867  
DB 200065 TTTGAAATCTTCTTATTATGTCCTTAATGAGTTTGAACCTTTAATTAATGATTAA 200118

RESULT 22  
US-09-754-853A-3  
Sequence 3, Application US/09754853A  
Publication No. US20030003491A1  
GENERAL INFORMATION:  
APPLICANT: Hauge, Brian M.  
APPLICANT: Parsons, Jeremy D.  
APPLICANT: Wang, Ming Li  
TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
FILE REFERENCE: 38-10(15810)B  
CURRENT APPLICATION NUMBER: US/09/754,853A  
CURRENT FILING DATE: 2001-01-05  
PRIOR APPLICATION NUMBER: US 60/174,880  
PRIOR FILING DATE: 2000-01-07  
NUMBER OF SEQ ID NOS: 1119  
SEQ ID NO 3  
LENGTH: 335913  
TYPE: DNA  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (46798)..(48763)..(48975)..(49573)  
OTHER INFORMATION: Clone ID: 240017\_region\_G3  
US-09-754-853A-3

Query Match 3.2%; Score 34.8; DB 9; Length 335913;  
Best Local Similarity 44.9%; Pred. No. 1.6e+02;

Matches 132; Conservative 0; Mismatches 162; Indels 0; Gaps 0;  
QY 574 AGACAAGTATTGATGGGAGCTCAAAATCCTGAACGGGATATAGGAAATGCTTTC 633  
DB 199825 AAACGACGATGCTTTGCTATAGGACAGTTGCCAATTGAAGATACAAAGTCTTGACC 199884  
QY 634 TAATTCGAAGAATTAAATGACTGCTTGAAGAAATCAGAAAGGAGCCGAAAGC 693  
DB 199885 TTAACCTTTGCAATTCATCATCTGCGAAATCAACGACTAAAGAAACCGTATGC 199944  
QY 694 TTCTAAGTAATCAAGAAATCTTGGACAATAGCTTTTCCATAGAGAAACCGTGAGGA 753  
DB 199945 ATTTAGAATGTAAATAATTTGGTGTCTATATCACTTAATTAAGGAAAG 200004  
QY 754 AAGGAGTCGATACCAATTAATTCGCAATAGCTTTTCCATAGAGAAACCGTGAGGA 813  
DB 200005 AGTGTACAGGTACAAACAATTTGAACCTTTTCAATTAATTAATGCGATAT 200064  
QY 814 GGAGCAACCTCCCTTTGATCTTTGATATGACAGAGATGACAAAGAGTAA 867  
DB 200065 TTTGAAATCTTCTTATTATGTCCTTAATGAGTTTGAACCTTTAATTAATGATTAA 200118

RESULT 23  
US-09-925-301-66  
Sequence 66, Application US/09925301  
Patent No. US20020052308A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA106  
CURRENT APPLICATION NUMBER: US/09/925,301  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05882  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1694  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 66  
LENGTH: 1391  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (16)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (20)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (25)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (27)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (1343)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (1356)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-925-301-66

Query Match 3.2%; Score 34.6; DB 10; Length 1391;  
Best Local Similarity 57.0%; Pred. No. 19;  
Matches 61; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 628 TGTTCCTAATTCGAAGAATTTAATGACTGCTTGTATGCTTCAAGAAATCAGAAAGGAGACC 687

```
Db 1253 TGTTCTATCAATAAAGTTATTCGGCTTCTTGTGATGAAAAAATTTTTTTT 1312
Qy 588 GAAAGCTTCTAAGTATCAAGAAAAATCTTGGACATAGCTTTTCC 734
Db 1313 AAAAAAATTTTTTTTAAAGGCGGNGCGCGAGGCTTTTTC 1359

RESULT 24
US-09-925-302-310/c
; Sequence 310, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 310
; LENGTH: 3092
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (3086)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (3089)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-310

Query Match
Best Local Similarity 3.2%; Score 34.6; DB 10; Length 3092;
Matches 82; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

Qy 29 TATTTCTCTGATTGTAAGGGGAATGATCTTAACAGACCTTGGTCGAGCATCTCT 88
Db 3027 TTTTAAATCAATAAACATGATGATTTATTTAAAGTCAGTTTGGTACATGATACAGATT 2968
Qy 89 AATTTGAAGTGAAGTCTACCAAGTTCATCAAAATGCTTATTTGAAATAAATCTCTTGT 148
Db 2967 GGTTTGCAGTTTAAATGAAGTGAATAGAAATGCTTAAATACAGCATCTCGCTGT 2908
Qy 149 GGAGTTGATGATGTAAGGGAAGAGCAATAATCTGATCCAA 189
Db 2907 GCAACAAATATCTGTAAAGTAAATAAGGTATTTGTATAGAA 2867

RESULT 25
US-09-764-864-1600
; Sequence 1600, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1600
; LENGTH: 10195
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
US-09-764-864-1600

Query Match
Best Local Similarity 3.2%; Score 34.4; DB 10; Length 10195;
Matches 86; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

Qy 874 GAAAAATCTAAATAGAAATGCAATTTGTGTGAGCTAAAGCTTGTCTCTTTATTTTC 933
Db 1884 GAAAAGAACAAAGATTTCTATCATTTAGATTGGGGTAAAAAGTTTGTGTGTTTTTTG 1943
Qy 934 CTTTGTAGAAATGATTCGGTAGTAGGAATATCGGTCGAGTATGCACCTTTGGAGTGAATG 993
Db 1944 TTTTGTGTTTTTTTTTTTAAAGATGGAGTTTACTCGTGTGCCAGGCTGGAGTCAAT 2003
Qy 994 GCTGCTCTGGAAGAACAGGAAAGTGAATGTCACAGCTTTAGAGCAATCTTC 1045
Db 2004 GCTCGATCTCAGCTCACTGCAACCTCTCTCCCGGTTCAAGGGATCTC 2055

RESULT 26
US-09-925-301-478
; Sequence 478, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 478
; LENGTH: 4201
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4077)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (4161)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (4186)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-478

Query Match
Best Local Similarity 3.1%; Score 34.2; DB 10; Length 4201;
Matches 63; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 834 TCTTTGATAATGACAGAAGATGAACAAAGAGTAAGTTTTTGAAAAATCTTAAATAAGAA 893
Db 2868 TTTTTTTTTTTAAACCATAGTTCATCGAAATGACGCTTTTCTGAACCTCAACATAGAA 2927
Qy 894 TGCATTTCTGTCGAGCTAAAGCTTGTCTCTTTATTTCTCTTTGTAGAA 944
Db 2928 CCCATTTTAACTGAAGTGAAGTGAAGCAAAATCATCTTTTTCATCTTTTAGGA 2978

RESULT 27
US-09-962-436-306
; Sequence 306, Application US/09962436
; Patent No. US20020081301A1
; GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-75
```



	Query Match Best Local Similarity	3.1%; 52.9%;	Score 34; DB 10;	Length 35641;
Matches	73; Conservative	0; Mismatches 65;	Indels 0;	Gaps 0;
Oy	612 GATAATAGGAAAATGTGGTTCTTAATTCCGAAGAATTAATGACTCGTTCGTCGTCAAGA	671		
Dd	1076 GTTAAATGACAAAGCTGGGATTTAATCAGGGATATTAATATCTCGTGCCGTFAAAAAGA	1135		
Oy	672 AATCAGAAGGAGACCAGCAAGCTTCTTAGTAAATCAAGAAAAAATCTTCGACAATPAGTTTT	731		
Dd	1136 ACTCCGTGTGATGAGAAATGTGTATATAGATAGAAAAAAAAGGACTAAATAAAAATAGAT	1195		
Oy	732 TCCATAGAGAAACCCTGG	749		
Dd	1196 TTCGTGATGTTAACAGTG	1213		

```

RESULT 30
US-10-003-806-9
? Sequence 9, Application US/10003806
? Patent No. US20020119929A1
? GENERAL INFORMATION:
? APPLICANT: Bishop, Colin E.
? APPLICANT: Agoulnik, Alexander I.
? APPLICANT: Zhu, Qichao
? TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
? FILE REFERENCE: P0206605U/10024824
? CURRENT APPLICATION NUMBER: US/10/003,806
? CURRENT FILING DATE: 2001-11-02
? PRIOR APPLICATION NUMBER: 60/245,872
? PRIOR FILING DATE: 2000-11-03
? NUMBER OF SEQ ID NOS: 14
? SOFTWARE: PatentIn version 3.1
? SEQ ID NO 9
? LENGTH: 180557
? TYPE: DNA
? ORGANISM: Human
US-10-003-806-9

```

	Query Match	3.1%;	Score 33.8;	DB 12;	Length 180557;
	Best Local Similarity	53.4%;	Pred. No. 2.3e+02;		
	Matches	71;	Mismatches	62;	Indels 0; Gaps
Qy	741	AAACCGTGAGAAAGAGTGTCGATACCATTAATTCGCATTAAAAAGGAGTGCTTACAA	800		
Db	26298	ACACAGTGGTGAGAGCATTTTATACGATCTGAAATGCAAAACAGGCATATTTCTTTA	26357		
Qy	801	CAGCCATATCAGTGGACGACACTCCCTTCGATCTTTTGATATATGACAGATGTAACAA	860		

Db 26358 TACTTTCAAAAGTGTAATACATACGTTGTACTTATCTGTCAGTGAAAAAGATTAAAAAC 26417

Qy 861 AGAGTAAGTTTTT 873

Db 26418 ACAAGAGTTTTT 26430

#### RESULT 31

US-09-742-312-3

; Sequence 3, Application US/09742312

; Patent No. US20020045166A1

; GENERAL INFORMATION:

; APPLICANT: CHANDRAMOULISARAN, Ishwar et al

; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,

; TITLE OF INVENTION: AND USES THEREOF

; FILE REFERENCE: CL000838

; CURRENT APPLICATION NUMBER: US/09/742.312

; CURRENT FILING DATE: 2000-12-22

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 147309

; TYPE: DNA

; ORGANISM: Human

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)...(147309)

; OTHER INFORMATION: n = A,T,C or G

US-09-742-312-3

Query Match 3.1%; Score 33.6; DB 10; Length 147309;

Best Local Similarity 48.0%; Pred. No. 2.4e+02;

Matches 96; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

Qy 694 TTCTAAGTCAATCAAGAAAAATCTTTGGACAATAGCTTTTCCATAGAGAAAAACCGTGAGGA 753

Db 35134 TTCAAGTCAACAATCCATATTTTTCAGACAACATGTTGTCTATGTAGAAAAATCCCAAAG 35193

Qy 754 AAGGAGTCGATACATAAATCCGATTTAAAAAGGGAGTGCTTACACAGCCATATCAGT 813

Db 35194 AATCAAGGCACACTAGTAGTGAAGTTAGCAAGTGCACAGATATAAGGACAACACACAA 35253

Qy 814 GGAGCACTTCCTCTTGATCTTTTGATAATGACAGAGATGAACAAGAGTAAGTTTTT 873

Db 35254 GAATCAACATTTCTATATATATGATGAATTAATGAAGAAAAATTTGTAATATTT 35313

Qy 874 GAAAAATCTAAATAGAA 893

Db 35314 CCAATAGCTCAAAAAATGAA 35333

#### RESULT 32

US-09-880-107-2219/c

; Sequence 2219, Application US/09880107

; Patent No. US20020142981A1

; GENERAL INFORMATION:

; APPLICANT: Horne, Darci T.

; APPLICANT: Vockley, Joseph G.

; APPLICANT: Scherf, Uwe

; APPLICANT: Gene Logic, Inc.

; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer

; FILE REFERENCE: 44921-5028-WO

; CURRENT APPLICATION NUMBER: US/09/880.107

; CURRENT FILING DATE: 2001-06-14

; PRIOR APPLICATION NUMBER: US 60/211,379

; PRIOR FILING DATE: 2000-06-14

; PRIOR APPLICATION NUMBER: US 60/237,054

; NUMBER OF SEQ ID NOS: 3950

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2219

; LENGTH: 2679

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: Genbank Accession No. US20020142981A1 L47726

US-09-880-107-2219

#### Query Match

Best Local Similarity 3.1%; Score 33.4; DB 10; Length 2679;

Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 180 TCTGATCCAATTTCTAAAGGGAGAACACGTACCTTATCAATGAAAGCGTGTCCCAAGTTCCT 239

Db 1586 TCTTCTCCAGCTCCAGGGGAGAAAGCTTTGGCTTCTCTGATAAGCAGTACTGTAATTTCAC 1527

Qy 240 GAGGAGAACTGATAAGAACGGCTCCATATGCTTTTCTCTTCATGTTTTCATTAAGTC 299

Db 1526 CAAAGGATGACAGAGGCCAGCACCATATGCTTTTATGGAGTCTCTTGTTCGAGAGCC 1467

Qy 300 CGCTTTTCAACAGTAAACCA 318

Db 1466 CAAACTCCACAGTAAACCA 1448

#### RESULT 33

US-09-782-378A-20/c

; Sequence 20, Application US/09782378A

; Patent No. US20020102751A1

; GENERAL INFORMATION:

; APPLICANT: Hearing, Patrick

; APPLICANT: Bahou, Wadie

; APPLICANT: Sandalon, Ziv

; APPLICANT: Gnatenko, Dmitri

; TITLE OF INVENTION: Adenoviral Vectors

; FILE REFERENCE: STONYB-04970

; CURRENT APPLICATION NUMBER: US/09/782.378A

; CURRENT FILING DATE: 2001-02-12

; PRIOR APPLICATION NUMBER: 60/237,747

; NUMBER OF SEQ ID NOS: 27

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 20

; LENGTH: 2680

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-782-378A-20

#### Query Match

Best Local Similarity 3.1%; Score 33.4; DB 10; Length 2680;

Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 180 TCTGATCCAATTTCTAAAGGGAGAACACGTACCTTATCAATGAAAGCGTGTCCCAAGTTCCT 239

Db 1586 TCTTCTCCAGCTCCAGGGGAGAAAGCTTTGGCTTCTCTGATAAGCAGTACTGTAATTTCAC 1527

Qy 240 GAGGAGAACTGATAAGAACGGCTCCATATGCTTTTCTCTTCATGTTTTCATTAAGTC 299

Db 1526 CAAAGGATGACAGAGGCCAGCACCATATGCTTTTATGGAGTCTCTTGTTCGAGAGCC 1467

Qy 300 CGCTTTTCAACAGTAAACCA 318

Db 1466 CAAACTCCACAGTAAACCA 1448

#### RESULT 34

US-10-109-551-1

; Sequence 1, Application US/10109551

; Publication No. US20020194635A1

; GENERAL INFORMATION:

; APPLICANT: DUNNE, PATRICK W.

; APPLICANT: PIEDRAHITA, JORGE

; TITLE OF INVENTION: TRANSGENIC ANIMALS RESISTANT TO TRANSMISSIBLE

; TITLE OF INVENTION: SPONGIFORM ENCEPHALOPATHIES

; FILE REFERENCE: TAMK:20705

```

; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 9781
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009266.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.4
; US-09-864-761-9781

Query Match          3.1%; Score 33.2; DB 10; Length 568;
Best Local Similarity 54.0%; Pred. No. 30;
Matches 68; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 864 GTAAGCTTTTGA AAAATCTCTAAATAGAAATGCAATTTGTGTCGAGCTAAAGCTTGCCTTCT 923
DB 87 GTAAGCTTTTACATATCTCAAAAATAGAAAAGGAATCAAAACAAGCTTAAATATTAAATAA 146

QY 924 CTTATATTTCTTTTGGAAGATGATTCGGTAGTAGGAATATGGGTCGAGTAGTACACGTT 983
DB 147 GAGAAATAAATAACCTATATAGACTTATATGTATGTATGTATGTGTGTATGTGTCGCTG 206

QY 984 GGAGTG 989
DB 207 TGTGTG 212

RESULT 36
US-09-764-864-1784/c
; Sequence 1784, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1784
; LENGTH: 1738
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-864-1784

Query Match          3.1%; Score 33.2; DB 10; Length 1738;
Best Local Similarity 50.6%; Pred. No. 49;
Matches 80; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 829 TTGATCTTTGATTAATCAGACAGATGACAAAGAGTAGTTTTCGAAAATCTTAAAT 888
DB 739 TCCATTCATAGATCCTCATTTAACTACGAAAAAATATATCTGGGAAAAGTTTATGAA 680

QY 889 AGAAATGCAATTTGTGTCGAGCTAAAGCTTGTCTCTCTTTATTTTCTTTGTAGATGAT 948
DB 679 TGTAGTGAATATGGGAAGACTTTTCAGCAATGATTCAGATGTTTATTTTCAGAGTTA 620

QY 949 TCGGTAGTAGGAATATGGGTCGAGTATGCACGTTTGA 986
DB 619 TACTGTAGAGAATCATATGAAGACCATAAATGTGGGA 582

```

```
RESULT 37
US-09-938-842A-4741
; Sequence 4741, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPT300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4741
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4741

Query Match      3.1%; Score 33.2; DB 9; Length 2000;
Best Local Similarity 49.4%; Pred. No. 51;
Matches 86; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

Qy 275 CTTCCCTTCATGGTTTCAATAAGTCGCTTTCAACAGTCAACAAAGCAGCGTACAATA 334
Db 356 CATAATTCACCAATTTCAAGAAAAGCTTTATATCATGTGATCTAATAAGGTTACAAT 415
Qy 335 GCGATCAGATTCCTGTGTAGGTTTGTATGGTTTCTTACTAGAGAGCTTGTACT 394
Db 416 GTTGAGACTTACATTTTGGAGAAGTATTCGTGATCTATTAGTAGTACTGATAT 475
Qy 395 TTTCTATACATTTAGTAGAGAGCTTCCCATGTTTATGAAATTCAGAAAT 448
Db 476 GTTGGCTTCTTTTCTGAAAGATCAGATATCTGCCACAAATATTTTGTATAT 529

RESULT 38
US-09-880-107-3814
; Sequence 3814, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3814
; LENGTH: 198285
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 X87344
US-09-880-107-3814

Query Match      3.1%; Score 33.2; DB 10; Length 198285;
Best Local Similarity 48.4%; Pred. No. 3.3e+02;

Matches 92; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

Qy 698 AAGTAATCAAGAAAAAATCTTGAGCAATAGCTTTTCCATAGAGAAAAACCGTGAGAAAGG 757
Db 149825 AAATAATTAATAAATAGATGGACTTCTAGCTAGACTAATAAAGAAAAAGAGAGAGG 149884

Qy 758 AGTCGATACCAATAAATTCGATTAAAAAAGGGAGTGCTTCAACAGCCATATCATGTGGAG 817
Db 149885 ATTCAATAAACACAAATAGGAATGACAAGGAGGTATTACCACTGACCCACAGAAATA 149944

Qy 818 CAACGCTTCCCTTTGATTCTTGTATATGACAGAGATGACAAGAGTAGTATTGTTTGA 877
Db 149945 CAGACAACCATCAGAGAAATATTATGATACCTTCTATGACATAAAGTAGAAAAATCTAGAAG 150004

Qy 878 AATTCTAAAA 887
Db 150005 AAATGGATAA 150014

RESULT 39
US-09-878-574-733
; Sequence 733, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 733
; LENGTH: 328
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: LIB3028-047-Q1-B1-A3
US-09-878-574-733

Query Match      3.0%; Score 32.8; DB 10; Length 328;
Best Local Similarity 58.0%; Pred. No. 30;
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Qy 837 TTGATAATGACAGAGATGAACAAAGAGTAAGTTTGTGAAAAATTTCTAAATAGAAATGC 896
Db 213 TTGATAATCTCTGAAGTTGAAGTATGACACGTTTGTGAGATGGCTTTCTAATAGACTAC 272

Qy 897 ATTTGTGTCGAGCTAAAGCTTGTCTTCTTTTATTTTCTT 936
Db 273 ACTTTTGTGCTAATACAGTTGATGCTTCCTTCTTCTT 312

RESULT 40
US-09-878-574-3709/C
; Sequence 3709, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 3709
; LENGTH: 406
```





[illegible]

US-09-522-334-26

	Query Match	3.08;	Score 32.4;	DB 9;	Length 4436;
	Best Local Similarity	49.4%;	Pred. No. 1.2e-02;		
	Matches 84;	Conservative 0;	Mismatches 82;	Indels 0;	Gaps 0;
QY	602	CCTGAACGGGATAATAGGAAAAATGTGTTTCTTAATTCGAGAAGATTTAATGACGTCTTGA	561		
Db	3669	CAACAACCAACAAACAAAGGTGAATAATTCATTTCTGGTGTCAAAACAGATGCTTTTGTGAGG	3728		
QY	662	TGGTCTAAGAAATCAGAAAGGAGACCGGAAGCTTCTAAGTAATCAAGAAAAAATCTTGGG	721		
Db	3729	TAAATATAACAAGAAACAAGAGAGAGAGGTACCTGAGTATTCAGTTTCATTAGATGA	3788		
QY	722	CAATAGCTTTTTCATAGAGAAAACCGTGAGGAAGGAGTCGATACATAA	771		
Db	3789	GAAATTGTAGTCAATGGTCCCAAGGCCTTTAAGTAACACAGCTTAACCCATCA	3838		

RESULT 50

```

RESULT: 30
US-09-754-853A-4
; Sequence 4, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Soybean Cyst Nematode Resistance
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 4
; LENGTH: 513509
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (111805)..(113968),(114684)..(115204)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(513509)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 3180I3_region_A3
US-09-754-853A-4

```

	Query Match	3.0%;	Score 32.4;	DB 9;	Length 513509;
	Best Local Similarity	53.1%;	Pred. No. 5.3e+02;		
	Matches 69;	Conservative 0;	Mismatches 61;	Indels 0;	Gaps 0;
QY	625	ATGTGTTTCTAATTCGAAGAATTTAATGACTGCTGTGATGCTTAAGAAATCAGAAAGGAG	684		
Db	263074	ATGAAATATATAATTTTGAATATATATATTTTATGTTTTCAGAAATAAAAAGCAAT	263133		
QY	685	ACCGAAAGCTTCTAAGTAATCAAGAAAAAATCTTTGGACAATAGCTTTTCCATAGACAAAA	744		
Db	263134	TTTTAAATGTTCTTATGAATTTTGAAAAAATATCTTTAGAAGAAATATTTTATATAAATG	263193		
QY	745	CCGTGAGGAA	754		
Db	263194	AATTGAGTAA	263203		

Search completed: January 9, 2003, 15:15:20  
Job time : 2105 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.  
OM nucleic - nucleic search, using sw model  
Run on: January 9, 2003, 06:43:27 ; Search time 39 Seconds  
(without alignments)  
8539.768 Million cell updates/sec

Title: US-09-438-185a-l\_COPY\_1199590\_1200675  
Perfect score: 1086  
Sequence: 1 ttggcaagctacctaacaac.....gtctctcgcagtagtgac 1086

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues  
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 150 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/PTCUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Description	
Result No.	Score	Query Match	Length DB ID		
C 1	72.4	6.7	7218	1	US-08-232-463-14
C 2	39.8	3.7	55827	4	US-09-813-133A-3
C 3	36.6	3.4	2098	4	US-08-937-931-5
C 4	36.6	3.4	2098	4	US-09-285-502-5
C 5	36.6	3.4	2098	4	US-09-709-126-5
C 6	36.6	3.4	2098	4	US-09-871-385A-5
C 7	36.4	3.4	1241	4	US-09-737-698B-25
C 8	34.8	3.2	6877	1	US-08-347-340-1
C 9	34.6	3.2	10952	1	US-08-602-036A-1
C 10	34.6	3.2	10952	2	US-08-642-407A-1
C 11	34.6	3.2	10952	2	US-08-502-374A-1
C 12	34	3.1	2067	4	US-09-106-194-11
C 13	33.8	3.1	65042	4	US-09-784-316-3
C 14	33.4	3.1	112132	4	US-08-938-546-1
C 15	33.4	3.1	1002	3	US-09-340-812-1
C 16	33.4	3.1	1020	3	US-08-938-546-3
C 17	33.4	3.1	1020	3	US-09-340-812-3
C 18	33.4	3.1	3360	3	US-08-938-546-5
C 19	33.4	3.1	3360	3	US-09-340-812-5
C 20	33.4	3.1	6029	4	US-09-149-727-1
C 21	33.4	3.1	10815	4	US-09-004-838-21
C 22	33	3.0	13149	4	US-09-004-838-87
C 23	32.8	3.0	806	4	US-08-936-165A-6
C 24	32.6	3.0	448	1	US-08-702-080-7
C 25	32.6	3.0	448	2	US-08-858-830-7
C 26	32.6	3.0	448	2	US-08-858-834-7
C 27	32.6	3.0	448	2	US-08-858-834-7

Sequence 16, Appl	US-08-930-285-16	3.0	638	4	Sequence 16, Appl
Sequence 3, Appl	US-08-484-438-3	3.0	5555	1	Sequence 3, Appl
Sequence 15, Appl	US-08-965-762-15	3.0	3800	4	Sequence 15, Appl
Sequence 3, Appl	US-09-851-896-3	3.0	70000	4	Sequence 3, Appl
Sequence 2094, Ap	US-09-134-001C-2094	3.0	528	4	Sequence 2094, Ap
Sequence 11, Appl	US-08-568-459A-11	2.9	8220	2	Sequence 11, Appl
Sequence 11, Appl	US-08-487-826B-11	2.9	8220	4	Sequence 11, Appl
Sequence 13, Appl	US-09-210-288-11	2.9	8220	4	Sequence 13, Appl
Sequence 12, Appl	US-08-487-826B-13	2.9	19124	2	Sequence 12, Appl
Sequence 3, Appl	US-09-356-952-12	2.9	43676	3	Sequence 3, Appl
Sequence 3, Appl	US-08-463-900B-3	2.9	1002	1	Sequence 3, Appl
Sequence 3, Appl	US-07-667-276A-1	2.9	1584	1	Sequence 3, Appl
Sequence 51, Appl	US-08-642-274D-51	2.9	679	4	Sequence 51, Appl
Sequence 51, Appl	US-08-952-014C-51	2.9	679	4	Sequence 51, Appl
Sequence 2, Appl	US-08-858-052-2	2.9	1734	2	Sequence 2, Appl
Sequence 2, Appl	US-09-200-284-2	2.9	1734	3	Sequence 2, Appl
Sequence 1, Appl	US-09-484-970B-60	2.9	13977	4	Sequence 1, Appl
Sequence 1, Appl	US-09-426-290-1	2.9	168575	4	Sequence 1, Appl
Sequence 116, App	US-09-428-034-1	2.9	1364	4	Sequence 116, App
Sequence 17, Appl	US-09-004-838-116	2.9	7521	4	Sequence 17, Appl
Sequence 17, Appl	US-09-007-005-17	2.9	289	4	Sequence 17, Appl
Sequence 1809, Ap	US-09-244-796-17	2.9	289	4	Sequence 1809, Ap
Sequence 42, Appl	US-09-134-001C-1809	2.9	999	4	Sequence 42, Appl
Sequence 15, Appl	US-09-018-635-42	2.9	591	4	Sequence 15, Appl
Sequence 52, Appl	US-08-697-826A-15	2.9	882	4	Sequence 52, Appl
Sequence 22, Appl	US-09-288-143-52	2.9	979	4	Sequence 22, Appl
Sequence 76, Appl	US-09-018-635-22	2.9	1641	4	Sequence 76, Appl
Sequence 1, Appl	US-08-961-527-76	2.9	10011	4	Sequence 1, Appl
Sequence 105, App	US-09-345-882-1	2.9	162450	4	Sequence 105, App
Sequence 105, App	US-08-413-118-105	2.8	2951	1	Sequence 105, App
Sequence 175, App	US-08-473-446-105	2.8	2951	3	Sequence 175, App
Sequence 1598, Ap	US-08-961-527-175	2.8	2951	4	Sequence 1598, Ap
Patent No. 5436393	US-09-134-001C-1598	2.8	1450	6	Patent No. 5436393
Sequence 8, Appl	5436393-2	2.8	1529	6	Sequence 8, Appl
Sequence 1, Appl	US-08-283-917-8	2.8	2085	2	Sequence 1, Appl
Sequence 1, Appl	US-08-961-716-8	2.8	2085	2	Sequence 1, Appl
Sequence 27, Appl	US-08-951-871-1	2.8	2739	1	Sequence 27, Appl
Sequence 27, Appl	US-08-431-080-27	2.8	4599	2	Sequence 27, Appl
Sequence 27, Appl	US-09-338-534-27	2.8	4599	2	Sequence 27, Appl
Sequence 1, Appl	US-09-345-234-27	2.8	4599	2	Sequence 1, Appl
Sequence 1, Appl	US-08-463-418-1	2.8	5613	3	Sequence 1, Appl
Sequence 10, Appl	US-08-947-823-10	2.8	51952	3	Sequence 10, Appl
Sequence 10, Appl	US-08-392-623-10	2.8	423	2	Sequence 10, Appl
Sequence 1741, Ap	US-08-466-961A-10	2.8	423	2	Sequence 1741, Ap
Sequence 21, Appl	US-09-134-001C-1741	2.8	915	4	Sequence 21, Appl
Sequence 21, Appl	US-08-484-105-21	2.8	1676	1	Sequence 21, Appl
Sequence 27, Appl	US-08-484-106-21	2.8	1676	1	Sequence 27, Appl
Sequence 16, Appl	US-08-627-907A-3	2.8	1942	3	Sequence 16, Appl
Sequence 16, Appl	US-09-219-983A-27	2.8	2092	4	Sequence 16, Appl
Sequence 18, Appl	US-08-392-623-16	2.8	8700	2	Sequence 18, Appl
Sequence 20, Appl	US-08-466-961A-16	2.8	8700	2	Sequence 20, Appl
Sequence 248, App	US-08-645-193B-18	2.8	475	2	Sequence 248, App
Sequence 4, Appl	US-08-623-906A-20	2.8	1302	4	Sequence 4, Appl
Sequence 3, Appl	US-09-134-001C-248	2.8	1437	4	Sequence 3, Appl
Sequence 3, Appl	US-09-137-223A-4	2.8	1880	2	Sequence 3, Appl
Sequence 3, Appl	US-08-845-161A-3	2.8	1880	4	Sequence 3, Appl
Sequence 3, Appl	US-09-270-751-3	2.8	1880	4	Sequence 3, Appl
Sequence 18, Appl	US-09-168-218B-2	2.8	2598	4	Sequence 18, Appl
Sequence 17, Appl	US-09-026-033-18	2.8	2602	4	Sequence 17, Appl
Sequence 272, App	US-08-838-151A-17	2.8	598	4	Sequence 272, App
Sequence 272, App	US-09-556-877-272	2.8	598	4	Sequence 272, App
Sequence 1070, Ap	US-09-620-412C-272	2.8	1316	4	Sequence 1070, Ap
Sequence 1, Appl	US-09-134-001C-1070	2.8	1316	4	Sequence 1, Appl
Sequence 2809, Ap	US-08-871-924A-1	2.8	1316	4	Sequence 2809, Ap
Sequence 2, Appl	US-09-361-900-1	2.8	1704	2	Sequence 2, Appl
Sequence 1, Appl	US-08-196-016A-1	2.8	2249	4	Sequence 1, Appl
Sequence 1, Appl	US-09-134-001C-2809	2.8	3496	2	Sequence 1, Appl
Sequence 195, App	US-08-974-546-2	2.8	3559	2	Sequence 195, App
Sequence 3, Appl	US-09-169-950-1	2.8	5372	3	Sequence 3, Appl
Sequence 3, Appl	US-09-070-060-1	2.8	6693	4	Sequence 3, Appl
Sequence 3, Appl	US-09-357-746-1	2.8	10607	1	Sequence 3, Appl
Sequence 3, Appl	US-08-961-527-195	2.8	10607	1	Sequence 3, Appl
Sequence 3, Appl	US-08-078-090-3	2.8	10607	1	Sequence 3, Appl



Db 908 TAAACTAAGTGTGTGTATGTAATGCTTTAAGGAGGCTTCCTTTCTTAACGATGG 849  
QY 1012 GGAAG 1017  
Db 848 GTGAGG 843

RESULT 2  
US-09-813-133A-3/C  
; Sequence 3, Application US/09813133A  
; Patent No. 6455294  
; GENERAL INFORMATION:  
; APPLICANT: GAN, Weiniu et al  
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: CL001173  
; CURRENT APPLICATION NUMBER: US/09/813/133A  
; CURRENT FILING DATE: 2001-06-06  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 55827  
; TYPE: DNA  
; ORGANISM: Human  
US-09-813-133A-3

Query Match 3.7%; Score 39.8; DB 4; Length 55827;  
Best Local Similarity 48.1%; Pred. No. 0.46; 122; Indels 0; Gaps 0;  
Matches 113; Conservative 0; Mismatches 0

QY 522 TATCGAGAGTTGCGATTACAGAGCGCAATGGGAAAGTAACGGTCTCGCAAGACAGCAAGT 581  
Db 29992 TCTCAAAAGCAGCTCAACAGACAGCAAAAATAGATAAAGGGGATTACATCAAACTGAAAA 29933  
QY 582 ATTGATGGGGAGCTACAAATCCTGAAACGGGATAATAGGAAAAATGTGTTCTTAATTGCA 641  
Db 29932 TTTGCTTCTGCAACAGCAAGGAAATACACGCTGAAAAAGAAAAATGGGAGAAATATG 29873  
QY 642 AGAATTTAATGACTGCTGTGATGGTCTAAGAAATCAGAAAGGAGACCGGAAAGCTTCTTAAGT 701  
Db 29872 TCGAAGCTATCCAGCAACAGGGACTTAATACAGAGATATATAGGAAGCGCAAAACAATT 29813  
QY 702 AATCAAGAAAAATCTTGGACAATAGCTTTTCATAGAGAAAAACCGTGAGGAAAG 756  
Db 29812 CAAAAGAAAAACAATAATCAATTTTAAAAATGAGAAAGGAATCTGAACAG 29758

RESULT 3  
US-08-937-931-5/C  
; Sequence 5, Application US/08937931  
; Patent No. 5935792  
; GENERAL INFORMATION:  
; APPLICANT: Rubin, Gerald M.  
; APPLICANT: Pan, Duojia  
; APPLICANT: Rooke, Jenny  
; APPLICANT: Yavari, Reza  
; APPLICANT: Xu, Tian  
; TITLE OF INVENTION: KUZ: A No. 5935792el Family of Metalloproteases  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
; STREET: 268 BUSH STREET, SUITE 3200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/937,931  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OSMAN, RICHARD A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: B97-081  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 343-4341  
; TELEFAX: (415) 343-4342  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2098 base pairs  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/937,931  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: OSMAN, RICHARD A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: B97-081  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 343-4341  
TELEFAX: (415) 343-4342  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2098 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-937-931-5

Query Match 3.4%; Score 36.6; DB 2; Length 2098;  
Best Local Similarity 65.1%; Pred. No. 0.87; 29; Indels 0; Gaps 0;  
Matches 54; Conservative 0; Mismatches 0

QY 816 AGCAACCTTCCCTTTGATTCTTGATAATGACAGAAGATGAACAAAGAGTAAGTTTTTGA 875  
Db 1136 AGCCACAAAAGCTTTGAAGGTTTGATATTTCAGAGAAGATGCCAAAGAGTCTCTTTATTT 1077

QY 876 AAAATCTAAATAGAAATGCAT 898  
Db 1076 AGGCTACTGAATATTATTAT 1054

RESULT 4  
US-09-285-502-5/C  
; Sequence 5, Application US/09285502  
; Patent No. 6190876  
; GENERAL INFORMATION:  
; APPLICANT: Rubin, Gerald M.  
; APPLICANT: Pan, Duojia  
; APPLICANT: Rooke, Jenny  
; APPLICANT: Yavari, Reza  
; APPLICANT: Xu, Tian  
; TITLE OF INVENTION: KUZ: A No. 6190876el Family of Metalloproteases  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
; STREET: 268 BUSH STREET, SUITE 3200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/937,931  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OSMAN, RICHARD A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: B97-081  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 343-4341  
; TELEFAX: (415) 343-4342  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2098 base pairs  
; CURRENT APPLICATION DATA:

; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
US-09-285-502-5

Query Match 3.4%; Score 36.6; DB 4; Length 2098;  
Best Local Similarity 65.1%; Pred. No. 0.87;  
Matches 54; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 816 AGCAACCTCCCTTGGATCTTTGATATGACAGAAAGATGACAAAGAGTAAGTTTGA 875

DB 1136 AGCCACAAAGCTTGAAGTTTGATATTCAGAGAAGATGACCAAGAGTCTCTTTATTT 1077

QY 876 AAAATCTTAATAGAAATGCAT 898

DB 1076 AGGCTACTGAAATATTATTAT 1054

## RESULT 5

US-09-709-126-5/c

; Sequence 5, Application US/09709126

; Patent No. 6319704

; GENERAL INFORMATION:

; APPLICANT: Rubin, Gerald M.

; Pan, Duojia

; Rooke, Jenny

; Yavari, Reza

; Xu, Tian

; TITLE OF INVENTION: KUZ: A No. 6319704el Family of Metalloproteases

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP

; STREET: 268 BUSH STREET, SUITE 3200

; CITY: SAN FRANCISCO

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/709,126

; FILING DATE: 08-No. 6319704-2000

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 09/285,502

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: OSMAN, RICHARD A

; REGISTRATION NUMBER: 36,627

; REFERENCE/DOCKET NUMBER: B97-081

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 343-4341

; TELEFAX: (415) 343-4342

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2098 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cdna

; SEQUENCE DESCRIPTION: SEQ ID NO: 5:

## Query Match

Best Local Similarity 3.4%; Score 36.6; DB 4; Length 2098;  
Matches 54; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 816 AGCAACCTCCCTTGGATCTTTGATATGACAGAAAGATGACAAAGAGTAAGTTTGA 875

DB 1136 AGCCACAAAGCTTGAAGTTTGATATTCAGAGAAGATGACCAAGAGTCTCTTTATTT 1077

DB 1136 AGCCACAAAGCTTGAAGTTTGATATTCAGAGAAGATGACCAAGAGTCTCTTTATTT 1077

QY 876 AAAATCTTAATAGAAATGCAT 898

DB 1076 AGGCTACTGAAATATTATTAT 1054

## RESULT 6

US-09-871-385A-5/c

; Sequence 5, Application US/09871385A

; Patent No. 6399350

; GENERAL INFORMATION:

; APPLICANT: Rubin, Gerald M.

; Pan, Duojia

; Rooke, Jenny

; Yavari, Reza

; Xu, Tian

; TITLE OF INVENTION: KUZ: A No. 6399350el Family of Metalloproteases

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP

; STREET: 268 BUSH STREET, SUITE 3200

; CITY: SAN FRANCISCO

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/871,385A

; FILING DATE: 31-May-2001

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/709,126

; FILING DATE: 08-No. 6399350-2000

; APPLICATION NUMBER: 09/285,502

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: OSMAN, RICHARD A

; REGISTRATION NUMBER: 36,627

; REFERENCE/DOCKET NUMBER: B97-081

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 343-4341

; TELEFAX: (415) 343-4342

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2098 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cdna

; SEQUENCE DESCRIPTION: SEQ ID NO: 5:

; US-09-871-385A-5

## Query Match

Best Local Similarity 3.4%; Score 36.6; DB 4; Length 2098;  
Matches 54; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 816 AGCAACCTCCCTTGGATCTTTGATATGACAGAAAGATGACAAAGAGTAAGTTTGA 875

DB 1136 AGCCACAAAGCTTGAAGTTTGATATTCAGAGAAGATGACCAAGAGTCTCTTTATTT 1077

QY 876 AAAATCTTAATAGAAATGCAT 898

DB 1076 AGGCTACTGAAATATTATTAT 1054

## RESULT 7

US-09-737-698B-25

; Sequence 25, Application US/09737698B

```
; Patent No. 6462258
; GENERAL INFORMATION:
; APPLICANT: Finch, Karen
; APPLICANT: Wilkinson, Jack
; TITLE OF INVENTION: No. 6462258el Plant Expression Constructs
; FILE REFERENCE: 38-21(51499)C
; CURRENT APPLICATION NUMBER: US/09/737,698B
; PRIOR FILING DATE: 2000-12-15
; PRIOR APPLICATION NUMBER: US 60/171,173
; PRIOR FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 25
; LENGTH: 1241
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1241)
; OTHER INFORMATION: n= a or g or c or t/u
; OTHER INFORMATION: Act7 promoter polynucleotide sequence and intron
; US-09-737-698B-25

Query Match          3.4%; Score 36.4; DB 4; Length 1241;
Best Local Similarity 59.8%; Pred. No. 0.79; Mismatches 0; Gaps 0;
Matches 61; Conservative 0;

QY 361 TAGCGTGTGTTTACTAGAGAGAGCTGTACTTTTCTCTATGACTTTAGTGAAGAGTCT 420
DB 1030 TAGGTGTGATGTTTACAGAGTGAATCTAGTCTTTCTCTTTGAGATCTGTGAAGTTGA 1089

QY 421 TCCATGTTTATGAAAAATTACAGAAATGAGGAGTGTAGAG 462
DB 1090 ACCTAGTTTCTCAATAATCAACATATGAAGCGATGTTTGAG 1131

RESULT 8
US-08-347-340-1
; Sequence 1, Application US/08347340
; Patent No. 5633439
; GENERAL INFORMATION:
; APPLICANT: WALTER, MICHAEL H
; TITLE OF INVENTION: EXPRESSION OF GENES IN TRANSGENIC PLANTS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: DC USA
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/347,340
; FILING DATE: 06-DEC-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/01098
; FILING DATE: 27-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9211416.4
; FILING DATE: 29-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N
; REGISTRATION NUMBER: 16773
; REFERENCE/DOCKET NUMBER: 9320/212448
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202/861-3000
; TELEFAX: 202/822-0944
```

```
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6877 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORGANISM: Nicotiana tabacum
; US-08-347-340-1

Query Match          3.2%; Score 34.8; DB 1; Length 6877;
Best Local Similarity 44.5%; Pred. No. 4.6; Mismatches 0; Gaps 0;
Matches 138; Conservative 0;

QY 592 AGCTACAAATCCCTGAAACGGGATAATAGGAAAAATGTGTTCTTAATTCGAGAATTTAAT 651
DB 61 ATCAATAAATGTCCTAACTGCAATGATCGTGTAAATTTGTGTATCAGAGAAATGAGTGACT 120
QY 652 GACTCGTGTGATGCTCTAAGAAATCAGAAAGGAGACCGGAAAGCTTCTTAAGTAATCAAGAAA 711
DB 121 GATATAGTAGGCAATTTGAAAGTTGAAAGATGAATTTACACTAAACTCAGATAAGATG 180
QY 712 AAATCTTTGGCAATAGCTTTTCCATAGAGAAAAACCGTGAGGAAAGGAGTCGATACATAA 771
DB 181 ATTTTGGGGTAGGCATTTGATTAATGCAAAACCTACTCTCATGAATGTCAGTTTAAC 240
QY 772 ATTCGGATTAATAAAGGAGTGTCTTACAAACAGCCATATCAGTGGAGCAACCTTCCCTTTG 831
DB 241 AACCAATCAACCAATGAATTTCTAAAAACAAGAAATTTACTGGGAAAAAACAACGTTTCG 300
QY 832 ATCTTTGTAATACAGAGAGATGAACAAAGAGTAAGTTTTTGAAAAAATCTAAATAAGA 891
DB 301 AATAATAAGAAATATCGTAAATTAATTAACCAACAATTTGCATTGGTTATAACTTA 360
QY 892 AATGCATTG 901
DB 361 TAATCACGTG 370

RESULT 9
US-08-602-036A-1
; Sequence 1, Application US/08602036A
; Patent No. 5789248
; GENERAL INFORMATION:
; APPLICANT: Oeystein, Fodstad
; APPLICANT: Hovig, Eivind
; APPLICANT: Engebraaten, Olav
; APPLICANT: Maelandsmo, Gunhild H.
; APPLICANT: Agrawal, Sudhir
; TITLE OF INVENTION: CAPL SPECIFIC OLIGONUCLEOTIDES AND
; TITLE OF INVENTION: METHODS OF INHIBITING METASTATIC CANCER
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HALE AND DORR LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: United States of America
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/602,036A
; FILING DATE: 16-FEB-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Kerner, Ann-Louise
; REGISTRATION NUMBER: 33,523
```

```

; REFERENCE/DOCKET NUMBER:  HYZ-039C1P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 526-6000
; TELEFAX: (617) 526-5000
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10952 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-602-036A-1

Query Match          3.2%   Score 34.
Best Local Similarity 55.4%  Pred. No.
Matches 67; Conservative 0; Mismatch

QY 410 GTCAAGAGCTCTCCCATGTTTATGAAAATTCA
      ||| ||| ||| ||| ||| |||
Db 7700 GTGAGGGAGACCCGCTGTTTCGACGACGATGA

QY 470 ACGTGCCCTAAAAGGTCATGGATGAGATCAGGA
      ||| ||| ||| ||| ||| |||
Db 7760 CTCGTCATAAAAGGTCCTTGGATGAGATCTGTT

QY 530 G 530
      |
Db 7820 G 7820

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-502-374A-1

Query Match 3.2%; Score 34.6; DB 2; Length 10952;
Best Local Similarity 55.4%; Pred. No. 6.5;
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps

QY 410 GTGAGAGCTCTTCCCATGTTTATGAAAAATCAGAAAATGAGGGATGTAGAAAGCCAAAGC 469
Db 7700 GTGGAGGAGACCCCGCTTTTTCGAGCAGCATGAGAAATATAGGCATTTCTCGAATCCACCGT 7759

QY 470 ACGTCCGCTAAAGGTCATGCGATGAGATCAGGAGTGAAGGAGAGAAATATCTTTATTCGAGA 529
Db 7760 CTCGCCATAAAGTCCCTGGATGAGATCTGTTATCAGTCAGGAATCAAGCTCTTCAGAGA 7819

QY 530 G 530
Db 7820 G 7820

RESULT 11
US-08-642-407A-1
; Sequence 1, Application US/08642407A
; Patent No. 5877308
; GENERAL INFORMATION:
; APPLICANT: Oeystein, Fodstad
; APPLICANT: Hovig, Eivind
; APPLICANT: Engebraaten, Olav
; APPLICANT: Maelandsmo, Gunhild H.
; APPLICANT: Agrawal, Sudhir
; TITLE OF INVENTION: CAPL SPECIFIC OLIGONUCLEOTIDES AND
; TITLE OF INVENTION: METHODS OF INHIBITING METASTATIC CANCER
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HALE AND DORR LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: United States of America
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642.407A
; FILING DATE: 03-May-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Kerner, Ann-Louise
; REGISTRATION NUMBER: 33,523
; REFERENCE/DOCKET NUMBER: Hyz-039CPDV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 526-6000
; TELEFAX: (617) 526-5000
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10952 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-642-407A-1

Query Match 3.2%; Score 34.6; DB 2; Length 10952;
Best Local Similarity 55.4%; Pred. No. 6.5;
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps

```







```

RESULT 17
US-08-938-546-3/c
: Sequence 3, Application US/08938546
: Patent No. 6004556
: GENERAL INFORMATION:
: APPLICANT: Fosberry, Andrew P.
: APPLICANT: Lawlor, Elizabeth J.
: APPLICANT: Nicholas, Richard O.
: TITLE OF INVENTION: No. 6004556el rsbu-1
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dechert Price & Rhoads
: STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
: CITY: Philadelphia
: STATE: PA
: COUNTRY: US
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/0938,546
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/938,546
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Dickinson, Todd Q
: REGISTRATION NUMBER: 28,354
: REFERENCE/DOCKET NUMBER: P50551
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 215-994-2252
: TELEFAX: 215-994-2252
: TELEX:
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1020 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
:
US-08-938-546-3
Query Match 3.1%; Score 33.4; DB 3; Length 1020;
Best Local Similarity 49.2%; Pred. No. 5;
Matches 88; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 865 TAAGTTTGTGAAATCTTAAATAGAAATGCAATTTGTCGAGCTAAAGCTTGCTCTC 924
DB 913 TAGTTTATGTTTTTAAATATATCTTAAAGTTTTTGTATCTATAAGGTCAC 854
QY 925 TTTATTTCTCTTGTAGATGATTCGGTAGTAGAATATGGGTCGAGTATGACGTTG 984
DB 853 TATTTCTAGCTTCAGTCACACCATCCGTTAAATGATTAATTAATCATCAAGTATAG 794
QY 985 GAGTGATTCGCTGCTTGGAGAACAGGAAAGTATGTCACGCTTTAGAGCAATCT 1043
DB 793 GAATTTCTTGTGATATCGTGTGTTGTAAGTATCTTAAACACTCTACCTCTAACT 735

RESULT 18
US-09-340-812-3/c
: Sequence 3, Application US/09340812
: Patent No. 6111078
: GENERAL INFORMATION:
: APPLICANT: Fosberry, Andrew P.
: APPLICANT: Lawlor, Elizabeth J.
: APPLICANT: Nicholas, Richard O.
: TITLE OF INVENTION: No. 6111078el rsbu-1
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dechert Price & Rhoads
: STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
: CITY: Philadelphia
: STATE: PA
: COUNTRY: US
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: 08/938,546
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/938,546
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Dickinson, Todd Q
: REGISTRATION NUMBER: 28,354
: REFERENCE/DOCKET NUMBER: P50551
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 215-994-2252
: TELEFAX: 215-994-2252
: TELEX:
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1020 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
:
US-09-340-812-3
Query Match 3.1%; Score 33.4; DB 3; Length 1020;
Best Local Similarity 49.2%; Pred. No. 5;
Matches 88; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 865 TAAGTTTGTGAAATCTTAAATAGAAATGCAATTTGTCGAGCTAAAGCTTGCTCTC 924
DB 913 TAGTTTATGTTTTTAAATATATCTTAAAGTTTTTGTATCTATAAGGTCAC 854
QY 925 TTTATTTCTCTTGTAGATGATTCGGTAGTAGAATATGGGTCGAGTATGACGTTG 984
DB 853 TATTTCTAGCTTCAGTCACACCATCCGTTAAATGATTAATTAATCATCAAGTATAG 794
QY 985 GAGTGATTCGCTGCTTGGAGAACAGGAAAGTATGTCACGCTTTAGAGCAATCT 1043
DB 793 GAATTTCTTGTGATATCGTGTGTTGTAAGTATCTTAAACACTCTACCTCTAACT 735

RESULT 19
US-08-938-546-5/c
: Sequence 5, Application US/08938546
: Patent No. 6004556
: GENERAL INFORMATION:
: APPLICANT: Fosberry, Andrew P.
: APPLICANT: Lawlor, Elizabeth J.
: APPLICANT: Nicholas, Richard O.
: TITLE OF INVENTION: No. 6004556el rsbu-1
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dechert Price & Rhoads
: STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
: CITY: Philadelphia
: STATE: PA
: COUNTRY: US
: ZIP: 19103
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: 08/938,546
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/938,546
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Dickinson, Todd Q
: REGISTRATION NUMBER: 28,354
: REFERENCE/DOCKET NUMBER: P50551
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 215-994-2252
: TELEFAX: 215-994-2252
: TELEX:
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1020 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
:
US-09-340-812-3
Query Match 3.1%; Score 33.4; DB 3; Length 1020;
Best Local Similarity 49.2%; Pred. No. 5;
Matches 88; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 865 TAAGTTTGTGAAATCTTAAATAGAAATGCAATTTGTCGAGCTAAAGCTTGCTCTC 924
DB 913 TAGTTTATGTTTTTAAATATATCTTAAAGTTTTTGTATCTATAAGGTCAC 854
QY 925 TTTATTTCTCTTGTAGATGATTCGGTAGTAGAATATGGGTCGAGTATGACGTTG 984
DB 853 TATTTCTAGCTTCAGTCACACCATCCGTTAAATGATTAATTAATCATCAAGTATAG 794
QY 985 GAGTGATTCGCTGCTTGGAGAACAGGAAAGTATGTCACGCTTTAGAGCAATCT 1043
DB 793 GAATTTCTTGTGATATCGTGTGTTGTAAGTATCTTAAACACTCTACCTCTAACT 735

```



us-09-438-185a-1\_1199590\_1200675.rni

Fri Jan 10 12:01:27 2003

```

GENERAL INFORMATION:
APPLICANT: Michelmore, Richard W.
APPLICANT: Shen, Kathy
APPLICANT: Meyers, Blake
TITLE OF INVENTION: Procedures and Materials for
TITLE OF INVENTION: Conferring Pest Resistance in Plants
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/004,838
FILING DATE: 09-JAN-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/781,734
FILING DATE: 10-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-0788100S
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 10815 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: -
LOCATION: 1..10815
OTHER INFORMATION: /note= "RLG2A"
US-09-004-838-21

```

```

Query Match 3.0%; Score 33; DB 4; Length 10815;
Best Local Similarity 48.6%; Pred. No. 18;
Matches 120; Conservative 0; Mismatches 125; Indels 2; Gaps 1;

QY 818 CAACCTTCCTTGGATCTTGTGATAATGACAGAGATGACAAAGAGTAAAGTTTGAAG 877
DB 7239 CAACATCCATTTGGATTTGGTGGCACTGGAGATGAATACACACAGTGGTGAAG 7298
QY 878 AATTCTAAAATAGAAATGCAATTTGTGCGAGCTTAAAGCTTGCTCTCTTTTCTTT 937
DB 7299 AATTATTTAAAGTGCATGTTGTGTAAGCTTTGTAATCT--CTTCCACACAATCCCATG 7356
QY 938 TGTAGAATGATTCGGTAGTAGGAATATGGGTCGAGATGCGAGCTTGGAGTATGGCTG 997
DB 7357 TCTATCTGATCATCTTGAAGAGCTTGAAGTCGAGAAATTTGGTTCATTGAATCGTTA 7416
QY 998 TTCTGGAAGAACAGGAAAGATGATTGTCTCAGCTTTAGACAACTTTCAGAAATFACITTT 1057
DB 7417 TTCAACATTGACTTGGATTGTGCTGCTGCAATTTGGGCAAGAGAACACAGCATCAGCTTA 7476
QY 1058 GGGTCCA 1064
DB 7477 AGAACA 7483

```

RESULT 23  
US-09-004-838-87

```

Sequence 87, Application US/09004838
Patent No. 6350933
GENERAL INFORMATION:
APPLICANT: Michelmore, Richard W.
APPLICANT: Shen, Kathy
APPLICANT: Meyers, Blake
TITLE OF INVENTION: Procedures and Materials for
TITLE OF INVENTION: Conferring Pest Resistance in Plants
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/004,838
FILING DATE: 09-JAN-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/781,734
FILING DATE: 10-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-0788100S
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 87:
SEQUENCE CHARACTERISTICS:
LENGTH: 13149 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: -
LOCATION: 1..13149
OTHER INFORMATION: /note= "RG2A"
US-09-004-838-87

Query Match 3.0%; Score 33; DB 4; Length 13149;
Best Local Similarity 48.6%; Pred. No. 20;
Matches 120; Conservative 0; Mismatches 125; Indels 2; Gaps 1;

QY 818 CAACCTTCCTTGGATCTTGTGATAATGACAGAGATGACAAAGAGTAAAGTTTGAAG 877
DB 7120 CAACATCCATTTGGATTTGGTGGCACTGGAGATGAATACACACAGTGGTGAAG 7179
QY 878 AATTCTAAAATAGAAATGCAATTTGTGCGAGCTTAAAGCTTGCTCTCTTTTCTTTT 937
DB 7180 AATTATTTAAAGTGCATGTTGTGTAAGCTTTGTAATCT--CTTCCACACAATCCCATG 7237
QY 938 TGTAGAATGATTCGGTAGTAGGAATATGGGTCGAGATGCGAGCTTGGAGTATGGCTG 997
DB 7238 TCTATCTGATCATCTTGAAGAGCTTGAAGTCGAGAAATTTGGTTCATTGAATCGTTA 7297
QY 998 TTCTGGAAGAACAGGAAAGATGATTGTCTCAGCTTTAGAGCAATCTTCAGAAATFACITTT 1057
DB 7298 TTCAACATTGACTTGGATTGTGCTGCTGCAATTTGGGCAAGAGAACACAGCATCAGCTTA 7357
QY 1058 GGGTCCA 1064
DB 7358 AGAACA 7364

```

RESULT 25  
US-08-702-080-7/c

ADDRESS: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140

ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/858,830

FILING DATE: 03/06/2007  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/702,080

APPLICATION NUMBER: 067/02/000  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWN, SCOTT A.  
REGISTRATION NUMBER: 32,724  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 448 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
06-08-858-830-7

Query Match	3.0%;	Score 32.6;	DB 2;	Length 448;
Best Local Similarity	60.9%;	Pred. No. 5.9;		
Matches 53:	Conservative	0;	Mismatches 34;	Indels 0;
				Gaps 0;

529	AGTTGCGCATACAGAGCGCAATCGGAAAGTACGCTCTGCAACAGACAGACAAGTATTGATG	588
530		
531		
532		
533		
534		
535		
536		
537		
538		
539		
540		
541		
542		
543		
544		
545		
546		
547		
548		
549		
550		
551		
552		
553		
554		
555		
556		
557		
558		
559	GGAGCTACAAATCCTGAACGGGATA	615
560		
561		
562		
563		
564		
565		
566		
567		
568		
569		
570		
571		
572		
573		
574		
575		
576		
577		
578		
579		
580		
581		
582		
583		
584		
585		
586		
587		
588		
589		
590		
591		
592		
593		
594		
595		
596		
597		
598		
599		
600		
601		
602		
603		
604		
605		
606		
607		
608		
609		
610		
611		
612		
613		
614		
615		
616		
617		
618		
619		
620		
621		
622		
623		
624		
625		
626		
627		
628		
629		
630		
631		
632		
633		
634		
635		
636		
637		
638		
639		
640		
641		
642	GGGTCTTTATCCCACTGCAACAGAAA	116
643		
644		
645		
646		
647		
648		
649		
650		
651		
652		
653		
654		

RESULT 27  
US-08-858-834-7/c  
; Sequence 7, Application US/08858834  
Patent No. 5081222

```

1 Patent No. 5981222
2
3 GENERAL INFORMATION:
4 APPLICANT: Jacobs, Kenneth
5 APPLICANT: McCoy, John
6 APPLICANT: LaVallie, Edward
7 APPLICANT: Racie, Lisa
8 APPLICANT: Meirberg, David
9 APPLICANT: Treacy, Maurice
10 APPLICANT: Spaulding, Vikki
11
12 TITLE OF INVENTION: SECRETED PROTEINS
13
14 TITLE OF INVENTION: ENCODING THEM
15
16 NUMBER OF SEQUENCES: 11
17
18 CORRESPONDENCE ADDRESS:
19 ADDRESSEE: Genetics Institute, Inc.
20 STREET: 87 CambridgePark Drive
21 CITY: Cambridge
22 STATE: Massachusetts
23 COUNTRY: U.S.A.
24
25 ZIP: 02140
26
27 COMPUTER READABLE FORM:
28 MEDIUM TYPE: Floppy disk
29 COMPUTER: IBM PC compatible
30 OPERATING SYSTEM: PC-DOS/MS-DOS
31 SOFTWARE: PatentIn Release #1.0, Ve

```

CURRENT APPLICATION DATA: US/08/858-834  
 APPLICATION NUMBER: US/08/858-834  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/702,080  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Brown, Scott A.  
 REGISTRATION NUMBER: 32,724  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (917) 438-8224  
 TELEFAX: (917) 438-8224  
 INFORMATION: F05 S0 176 851 7  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 448 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cdna  
 US-08-858-834-7

Query Match 3.0%; Score 32.6; DB 2; Length 448;  
Best Local Similarity 60.9%; Pred. No. 5.9;  
Matches 53; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

529	AGTTCGCATTACAGAGGCATGGGAAAGTACAGGTCTTCCACACAGACAGACAAGTATTCATG	588
QY		
202	AGGCTCATGACAGACAGACCATAGAAATCCCTGTTGTCGATCGGCATCAGGCCCTTGAGG	143
Db		
589	GGGAGCTACAAATCTCGAAACGGGATA	615
QY		
142	GGGCTTATCCCACTGCACAGGAA	116
Db		

```

RESULT 28
US-08 930-285-16
; Sequence 16, Application US/08930285
; Patent No. 6222099
; GENERAL INFORMATION:
; APPLICANT: Regents of the University
; TITLE OF INVENTION: TRANSGENIC PLANTS
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg, Wo
; STREET: P. O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/930,285
; FILING DATE: 13-APR-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/04625
; FILING DATE: 04-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Embretson, Janet E.
; REGISTRATION NUMBER: 39,665
; REFERENCE/DOCKET NUMBER: 600.3180US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 638 base pairs

```

; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE:  
; ORIGINAL SOURCE:  
US-08-930-285-16

Query Match 3.0%; Score 32.6; DB 4; Length 638;  
Best Local Similarity 52.6%; Pred. No. 6.9;  
Matches 71; Conservative 0; Mismatches 64; Indels 0; Gaps 0;  
QY 837 TTGATATGACAGACATCAACAGAGTAACTTTTCAAAAATTTCTAAATAGAAATGCC 896  
DB 190 TTGAATATGACATGCTTGACGAGATATTAACAGTTTTTTTATAAAAAACATGCATTT 249  
QY 897 ATTGTGTCAGCTAAGCTTCTCTCTTTTATTTTCTTTTCTAGAGATGTCGGTAGT 956  
DB 250 CTAGGAGTTGGACTAAGCTTTCTTTAGTATGAAGTGCCATGTTTTACATGGTCCATTTGT 309  
QY 957 AGGAATATGGGTGCG 971  
DB 310 GTCAATTTACAGTCG 324

RESULT 29  
US-08-484-438-3  
; Sequence 3, Application US/08484438  
; Patent No. 5811098  
; Patent No. 5811098 5780031  
; GENERAL INFORMATION:  
; APPLICANT: Plowman, Gregory D.  
; APPLICANT: Calouscou, Jean-Michel  
; APPLICANT: Shoyab, Mohammed  
; APPLICANT: Stegall, Clay B.  
; APPLICANT: Hellistr m, Ingeyerd  
; APPLICANT: Hellistr m, Karl E.  
; TITLE OF INVENTION: HER4 HUMAN RECEPTOR TYROSINE KINASE  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/484,438  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/323,442  
; FILING DATE: 14-OCT-1994  
; APPLICATION NUMBER: US 08/150,704  
; FILING DATE: 10-NOV-1993  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/981,165  
; FILING DATE: 24-NOV-1992  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Misrock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 5624-230  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5555 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 34..3210  
US-08-484-438-3

Query Match 3.0%; Score 32.4; DB 1; Length 5555;  
Best Local Similarity 53.1%; Pred. No. 20;  
Matches 69; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
QY 585 GATGGGAGCTACAATCTCTGAACGGGATAATAGGAAAAATGTTTCTTAATTCGAAGA 644  
DB 5423 GTTTGGTGGCTAATAACGTGCTAAGGGAACATCTTAAAAAGTGAATTTTGTATCAATAT 5482  
QY 645 ATTTAATGACTGCTTGTATGCTCTAAGAAATCAGAAAGGAGACCAGAAAGCTTCTTAAGTAAT 704  
DB 5483 TTCTTAAGCATATGCTATAGACTTTTGAACACCAAAAAAAGAAAAAAGAAAAA 5542  
QY 705 CAAGAAAAAA 714  
DB 5543 AAAAAA 5552

RESULT 30  
US-08-965-762-15  
; Sequence 15, Application US/08965762  
; Patent No. 6280963  
; GENERAL INFORMATION:  
; APPLICANT: Koltin, Yigal  
; APPLICANT: Gavrilas, Victoria  
; TITLE OF INVENTION: ESSENTIAL FUNGAL GENES AND THEIR USE  
; FILE REFERENCE: 07334/062001  
; CURRENT APPLICATION NUMBER: US/08/965,762  
; CURRENT FILING DATE: 1997-11-07  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 15  
; LENGTH: 3800  
; TYPE: DNA  
; ORGANISM: Saccharomyces cerevisiae  
US-08-965-762-15

Query Match 3.0%; Score 32.2; DB 4; Length 3800;  
Best Local Similarity 61.2%; Pred. No. 19;  
Matches 52; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
QY 856 AACAAAGAGTAAGTTTTTGAATAATCTAAATAGAAATGCAATTTGTCTGAGCTAAAGC 915  
DB 3 AAAGACAGTCAGACTTTTAAAGAAAGTAAACCAAAAGCTTTTTTTAGGACGGATTATACC 62  
QY 916 TTGCTTCTCTTTATTTTCTCTTTTGT 940  
DB 63 ATAGTCTCTCTTATTTCTCTTTT 87

RESULT 31  
US-09-851-896-3/c  
; Sequence 3, Application US/09851896  
; Patent No. 6410325  
; GENERAL INFORMATION:  
; APPLICANT: C. Frank Bennett  
; APPLICANT: Susan M. Freier  
; APPLICANT: Andrew T. Watt  
; TITLE OF INVENTION: ANTISENSE MODULATION OF PHOSPHOLIPASE A2, GROUP VI (CA2+-INDEP  
; TELECOMMUNICATION INFORMATION: EXPRESSION

```

; FILE REFERENCE: RFS-0220
; CURRENT APPLICATION NUMBER: US/09/851.896
; CURRENT FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 3
; TYPE: DNA
; LENGTH: 70000
; ORGANISM: Homo sapiens
; FEATURE:
; US-09-851-896-3

Query Match          3.0%; Score 32.2; DB 4; Length 70000;
Best Local Similarity 48.1%; Pred. No. 67;
Matches 91; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

Qy 727 GCTTTTCATAGAGAAACCGTGAGGAAAGAGTCGATACCATAAATTCGATTAATAAAA 786
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1889 GCTCCAGCGCTGGCAACAGAGTGAGACCCCTGCTCAAAACCAACCAACCAACCAA 1830
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 787 GGGAGTGCTTACACAGCCCATATCAGTGGGACCACTTCCTCTTGATTCCTTGATATGA 846
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1829 AAAAAACCCCAAAATCCCTGACATCCTGGAACTATATTTAGACAGAGCTACTAATA 1770
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 847 CAGAAGATGAACAGAGTAAGTTTGTGAAATCTTAAATAGAAATGCAATTTGTGTGCG 906
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1769 ATAAAGAAATGAACAGGTAATATTTACAAAGTTTATAGATGACAAATTAACACTAGG 1710
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 907 AGCTAAAGC 915
      || ||| |||
Db 1709 AGAAAAACC 1701

RESULT 32
; US-09-134-001C-2094
; Sequence 2094, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2094
; LENGTH: 528
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-2094

Query Match          2.9%; Score 32; DB 4; Length 528;
Best Local Similarity 55.4%; Pred. No. 9.3;
Matches 82; Conservative 0; Mismatches 65; Indels 1; Gaps 1;

Qy 767 CATAAATTCGATTAATAAGGAGTGCTTACACAGCCCATATCAGTGA-GCAACCTTC 825
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 305 CAGAAGACCAATAAATAAGAAACAGATTTTATATGCTATCAATCAACCCCTTTTG 364
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 826 CCTTTGATTCCTTTGATTAATGACAGAGATGAACAAAGAGTAAGTTTGTGAAATCTAA 885
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 365 TATCTCTTTATGACCAATTTGAAGGAGATAAAGATGTTAGGTGTTTAAATCTGTAA 424
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 886 AATAGAAATGCAATTTGTGTCGAGCTAAA 913
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 425 ATGATGAATCAATTAATATGGAAGTTAA 452
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 33
; US-08-568-459A-11
; Sequence 11, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitois, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; US-08-568-459A-11

; Sequence 11, Application US/08568459A
; Patent No. 5849306
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitois, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; US-08-568-459A-11

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/568.459A
; FILING DATE: 07-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8220 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-08-568-459A-11

Query Match          2.9%; Score 32; DB 2; Length 8220;
Best Local Similarity 56.7%; Pred. No. 31;
Matches 59; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 155 GATGTATTGAAGGGAAGACGAATAATCTGATCCAATTCTTAAGGGAGACAGTACGTTA 214
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2789 GGTGTCGCATGAGATAGGAACGGAATGTTCAATATTGAAGGAGAAACAAACACGTC 2848
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 215 TCAATGAACGCTGTCCAAAGTTCTCTGAGGAGAACTGATAAGAAC 258
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2849 TACAAAAAGCTCTTTTACCTCCCGACGAGAACACATGTGTAC 2892
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 34
; US-08-487-826B-11
; Sequence 11, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitois, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; US-08-487-826B-11
```





us-09-438-185a-l\_1199590\_1200675.rni

Fri Jan 10 12:01:27 2003

TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13

Query Match 2.9%; Score 32; DB 2; Length 19124;  
Best Local Similarity 56.7%; Pred. No. 44;  
Matches 59; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 155 GATGTATTGAAGGGAACGAATAATCTCATCAATCTTAAGGGGAGAACAGCTAGTTA 214  
DB 10109 GGTGTCCGATGAGTAATGAGCAAGGATGGTCAATATTGAAGGAAAAAACAACAGCTCA 10168

QY 215 TCAATGAAGCGTGTCCAAGTTCTCGAGGAGAACTGATAAGAAC 258  
DB 10169 TACAAAACGCTCTTTTACCTCCCGACGAGAACATGTGTAC 10212

RESULT 37  
US-09-356-952-12/c  
Sequence 12, Application US/09356952  
Patent No. 6117663  
GENERAL INFORMATION:  
APPLICANT: Borlack-Sjodin, Ann  
APPLICANT: Margarit, S. M.  
APPLICANT: Bor-Sogi, Dafna  
APPLICANT: Cole, Philip  
APPLICANT: Kuriyan, John  
TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE  
FILE REFERENCE: 600-1-228N  
CURRENT APPLICATION NUMBER: US/09/356,952  
CURRENT FILING DATE: 1999-07-19  
EARLIER APPLICATION NUMBER: 60/093,631  
EARLIER FILING DATE: 1998-07-21  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 12  
LENGTH: 43676  
TYPE: DNA  
ORGANISM: Saccharomyces cerevisiae  
US-09-356-952-12

Query Match 2.9%; Score 32; DB 3; Length 43676;  
Best Local Similarity 53.1%; Pred. No. 63;  
Matches 68; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 771 AATTCCGATTAAAAAGGAGTGTACACAGCCATATCAGTGGAGCAACCTCCCTTT 830  
DB 32382 AATTGAAAGTGAATAAACCTTTTCAGTTCATCTTCCAAATATTGGAGATCTCATCGCGTA 32323

QY 831 GATTCCTTTGATAATGACAGAGATGAACAAAGATGAAGTTTGTGAAAAATCTTAAATAG 890  
DB 32322 AAGGATTGATGCGCTCTTGAATGAAGAAATATTGACATAGTTTTTGAATAAATTTGTAANAAC 32263

QY 891 AATGCGAT 898  
DB 32262 TTCTGGAT 32255

RESULT 38  
US-08-463-090B-3  
Sequence 3, Application US/08463090B  
Patent No. 5801015  
GENERAL INFORMATION:

APPLICANT: Cottarel, Guillaume  
APPLICANT: Damagnez, Veronique  
APPLICANT: Draetta, Guilio  
TITLE OF INVENTION: Cell-Cycle Regulatory Proteins from  
TITLE OF INVENTION: Human Pathogens, and Uses Related Thereto  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley, Hoag & Eliot, LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII (text)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/463,090B  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Vincent, Matthew P.  
REGISTRATION NUMBER: 36,709  
REFERENCE/DOCKET NUMBER: MIV032.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 832-7000  
TELEFAX: (617) 832-7000  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1002 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 43..993  
US-08-463-090B-3

Query Match 2.9%; Score 31.8; DB 1; Length 1002;  
Best Local Similarity 47.3%; Pred. No. 14;  
Matches 96; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 87 CTAATTTTGAAGTGAGTTCTACAGTTTCATCAAAATCTTTATTGAAAAATAAAGTCTCTT 146  
DB 362 CTAATATGATAAAAAGATTTATGAATCAATTTATTCGAGGATTTAAACATGTCATCTC 421

QY 147 GTGGAGTTGATGATTTGAAGGAGAGACGAATATCTGATCCAAATCTTAAGGGAGAACAC 206  
DB 422 ATCGAGTTTACATCGTGATTTAAAACCAAAATTTATTGATTGATAAAGAAAGGAATT 481

QY 207 GTACGTTTATCAATGAAAGCGTGTCCTGAGGAGAACTGATGAAGACGGCTCCAT 266  
DB 482 TAAATATTAGCAGATTTGGATTAGCTCGAGCATTTGGAGTTCCATTAAAGACATATATCTC 541

QY 267 ATGCTTTTCTTCCTTCATGGTTT 289  
DB 542 ATGAACCTTGTCACCTTTATGGTAT 564

RESULT 39  
US-07-667-276A-1/c  
Sequence 1, Application US/07667276A  
Patent No. 5470971  
GENERAL INFORMATION:  
APPLICANT: Kondo, Keiji  
APPLICANT: Inouye, Masayori  
TITLE OF INVENTION: STRESS-INDUCED PROTEINS, GENES CODING  
TITLE OF INVENTION: THEREFOR, TRANSFORMED CELLS OF ORGANISMS, METHODS AND  
TITLE OF INVENTION: APPLICATIONS  
NUMBER OF SEQUENCES: 10





; Sequence 60, Application US/09484970B  
; Patent No. 6426186  
; GENERAL INFORMATION:  
; APPLICANT: Jones, Karen A.  
; APPLICANT: Volkmuth, Wayne  
; APPLICANT: Walker, Michael G.  
; TITLE OF INVENTION: BONE REMODELING GENES  
; FILE REFERENCE: PB-0014 US  
; CURRENT APPLICATION NUMBER: US/09/484,970B  
; CURRENT FILING DATE: 2000-01-18  
; NUMBER OF SEQ ID NOS: 172  
; SOFTWARE: PERL Program  
; SEQ ID NO: 60  
; LENGTH: 13977  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. 6426186 229357.11CB1  
; NAME/KEY: unsure  
; LOCATION: 11721-11761, 12294, 13969  
; OTHER INFORMATION: a, t, c, g, or other  
US-09-484-970B-60

Query Match 2.9%; Score 31.6; DB 4; Length 13977;  
Best Local Similarity 46.1%; Pred. No. 50;  
Matches 106; Conservative 0; Mismatches 124; Indels 0; Gaps 0;  
QY 765 ACCATAAATCCGATTAAAGGAGTGTACACAGCCATATCAGTGGAGCAACCTT 824  
DB 11867 ACACCAATGACTACTACACACTGCTCATTTGAGAACTGTGCTGAGTGGGCGGCTT 11926  
QY 825 CCCTTTCATCTTTGATATGACAGAAAGAGTAAAGTTTGAAGAAATCTA 884  
DB 11927 GAGTTTCATTCATATCTATATGCTATAGTATATATAAATCTATATAGAT 11986  
QY 885 AATAGAAATGCAATTTGTCGAGCTAAAGCTTCTCTCTTTATTTTCTTTGTAGAA 944  
DB 11987 AAGAGATAGCAATTTCTATAGACTGACTTTTCCATTTTAAATGTTCAATGCACATC 12046  
QY 945 TGATCGGTAGTAGGAATATGGGTGCGAGTAGCACTGTCAGCTGATGTTGG 994  
DB 12047 CTAATAGAAAGAAATTAATCTTAGTACGTACATCCAGGCTTACCTGCTTGG 12096

RESULT 45  
US-09-426-290-1  
; Sequence 1, Application US/09426290  
; Patent No. 6410712  
; GENERAL INFORMATION:  
; APPLICANT: Berglind Ran Olafsdottir  
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE  
; FILE REFERENCE: 2345, 2001-000  
; CURRENT APPLICATION NUMBER: US/09/426,290  
; CURRENT FILING DATE: 1999-10-25  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 168575  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (21181)...(21403)  
; NAME/KEY: CDS  
; LOCATION: (95252)...(95430)  
; NAME/KEY: CDS  
; LOCATION: (101753)...(101996)  
; NAME/KEY: CDS  
; LOCATION: (110324)...(110439)  
; NAME/KEY: CDS  
; LOCATION: (124058)...(124278)

; NAME/KEY: CDS  
; LOCATION: (127009)...(127130)  
; NAME/KEY: CDS  
; LOCATION: (128910)...(129139)  
US-09-426-290-1  
Query Match 2.9%; Score 31.6; DB 4; Length 168575;  
Best Local Similarity 50.0%; Pred. No. 1,4e+02;  
Matches 105; Conservative 0; Mismatches 104; Indels 1; Gaps 1;  
QY 684 GACCGAAGCTTCTAAGTAATCAAGAAAAAATCTTGGACAATAGCTTTTCCATAGAGAAA 743  
DB 78118 GAAGAAGCCAGATTTTATAAAGTAAAAATCTGCTTGTGCTCTTTCCA-AATTAGA 78176  
QY 744 ACCGTGAGGAAAGGAGTCGATACCATATAATTCGGATTAAAAAAGGAGTGTTCACACAG 803  
DB 78177 AGAGAAATGTATCATCTTAATACAGCAGATTCAAGTTATATAAAGACCTACTCCATCCAA 78236  
QY 804 CCATATCAGTGGAGCAACCTTCCCTTTTGATTTCTTTGATTAATGACAGAGATGAACACAGA 863  
DB 78237 AAAATTGAGTGAATAAAAAAGAAATTCGACTTACTTGTAAAGAGAAAAAGATTGCCAAGGC 78296  
QY 864 GTAAGTTTTTGAATAATCTTAAATAGAAA 893  
DB 78297 TTGCAGACTTGTGAGGTGTTAAATAACAA 78326

RESULT 46  
US-09-428-034-1/c  
; Sequence 1, Application US/09428034  
; Patent No. 6428996  
; GENERAL INFORMATION:  
; APPLICANT: Cheng, Kuo-Joan  
; APPLICANT: Lio, Jin-Hao  
; APPLICANT: Tsai, Cheng-Fang  
; APPLICANT: Hsu, Yih-Chin  
; TITLE OF INVENTION: CELLULOSE ENZYMES  
; FILE REFERENCE: 08919/036001  
; CURRENT APPLICATION NUMBER: US/09/428,034  
; CURRENT FILING DATE: 1999-10-27  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1364  
; TYPE: DNA  
; ORGANISM: Piromyces rhizinflata  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1128)  
US-09-428-034-1

Query Match 2.9%; Score 31.4; DB 4; Length 1364;  
Best Local Similarity 52.7%; Pred. No. 21;  
Matches 68; Conservative 0; Mismatches 51; Indels 0; Gaps 0;  
QY 829 TTGATCTTTGATTAATGACAGAGATGAACAAGAGTAAGTTTGGAAAAATCTTAAAT 888  
DB 1241 TTTTGTGATATATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1182  
QY 889 AGAAATGCAATTTGTGCGAGCTAAAGCTTCTCTCTTTATTTTCTCTTTTGTAGAATGAT 948  
DB 1181 ATTATATTTTATATATATATGTAATAATAATTTTTTTTTTTTAAATTTATTAGATGTT 1122  
QY 949 TCGGTAGTA 957  
DB 1121 GGATTAGCA 1113

RESULT 47  
US-09-004-838-116  
; Sequence 116, Application US/09004838  
; Patent No. 6350933  
; GENERAL INFORMATION:

APPLICANT: Michelmore, Richard W.  
APPLICANT: Shen, Kathy  
APPLICANT: Meyers, Blake  
TITLE OF INVENTION: Procedures and Materials for  
TITLE OF INVENTION: Confering Pest Resistance in Plants  
NUMBER OF SEQUENCES: 140  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/004,838  
FILING DATE: 09-JAN-1998  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/781,734  
FILING DATE: 10-JAN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Einhorn, Gregory P.  
REGISTRATION NUMBER: 38,440  
REFERENCE/DOCKET NUMBER: 023070-078810US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 116:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7521 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: -  
LOCATION: 1..7521  
OTHER INFORMATION: /note= "RG2N"  
US-09-004-838-116  
Query Match 2.9%; Score 31.4; DB 4; Length 7521;  
Best Local Similarity 48.5%; Pred. No. 43;  
Matches 116; Conservative 0; Mismatches 121; Indels 2; Gaps 1;  
QY 818 CAACCTTCCTTGGATTTGATTTGATTAAGACAGAGATGAACAAAGAGTAAGTTTTCGAA 877  
DB 4330 CACATCCATTTGGATTTGGTGTGCTGCACTGAGATGATACACACACAGTGTGAAG 4389  
QY 878 AATCTAAATAGAAATGCAATTTGTGCGAGCTAAAGCTTGCTTTCTTTTATTTTCCTTT 937  
DB 4390 AATATAAAGTGAATGATGTTGATGAATCT--CTTCCACAAATCCCATG 4447  
QY 938 TGTAGATGATTCGGTGTAGTATAGATATAGGGTGTGATGCGCTTGGAGTGTGCTG 997  
DB 4448 TCTATATGATCATCTTTGAAGAGCTTGAAGTGTGAGTGTGGTTCATTTGAATCGTTA 4507  
QY 998 TTCTGGAAGAACAGGGAAGTGTGCTGAGCTTTAGAGCAATCTTCAGATATATCTT 1056  
DB 4508 TTCAACATTTGACTTGGATTTGCTGCTCAATTTGGGCAAGAACAAAGAGAGCTT 4566  
RESULT 48  
US-09-007-005-17  
; Sequence 17, Application US/09007005B  
; Patent No. 6258558  
; GENERAL INFORMATION:  
; APPLICANT: Szostak, Jack W.  
; APPLICANT: Roberts, Richard W.  
; APPLICANT: Liu, Rihe  
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN  
; FILE REFERENCE: 00786/350007  
; CURRENT APPLICATION NUMBER: US/09/244,796  
; CURRENT FILING DATE: 1998-02-05  
; EARLIER APPLICATION NUMBER: 60/035,963  
; EARLIER FILING DATE: 1997-01-27  
; EARLIER APPLICATION NUMBER: 60/064,491  
; EARLIER FILING DATE: 1997-11-06  
; EARLIER APPLICATION NUMBER: 09/007,005  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17  
; LENGTH: 289  
; TYPE: RNA

APPLICANT: Liu, Rihe  
TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN  
FILE REFERENCE: 00786/350003  
CURRENT APPLICATION NUMBER: US/09/007,005B  
CURRENT FILING DATE: 1998-01-14  
EARLIER APPLICATION NUMBER: 60/035,963  
EARLIER FILING DATE: 1997-01-27  
EARLIER APPLICATION NUMBER: 60/064,491  
EARLIER FILING DATE: 1997-11-06  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 17  
LENGTH: 289  
TYPE: RNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Translation template  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(289)  
OTHER INFORMATION: n = A,T,C or G  
US-09-007-005-17  
Query Match 2.9%; Score 31.2; DB 4; Length 289;  
Best Local Similarity 9.3%; Pred. No. 12;  
Matches 26; Conservative 111; Mismatches 142; Indels 1; Gaps 1;  
QY 508 GGAGAAATATCTTTATCGAGAGTTCGCAATACAGAGCAATGGAAAGTAAGCTCTCTG 567  
DB 3 RGRGRCARARURURARURURURURURURURURURURURURURURURURURURURUR 62  
QY 568 CAACAGACAGACAGTATGTGAGGGAGCTACAATCTCGAAAGGATATAGGAAAAATG 627  
DB 63 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 122  
QY 628 TGTTCCTTAATTCGGAAGATTTAATGACTGCTGTGATGCTTAAGAAATCAACAAAGGAGACC 687  
DB 123 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 182  
QY 688 GAAAGCTTCTAAGTAATCAAGAA-AAAATCTTGGACAATACCTTTCCATAGAGAAACC 746  
DB 183 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 242  
QY 747 GTGAGGAAAGAGTCGATACCAATAAATTCGATTAAAAA 786  
DB 243 RURCRURURGRRCRUAIAAAAAAAAAAAAAAAAAAAAA 282  
RESULT 49  
US-09-244-796-17  
; Sequence 17, Application US/09244796  
; Patent No. 6281344  
; GENERAL INFORMATION:  
; APPLICANT: Szostak, Jack W.  
; APPLICANT: Roberts, Richard W.  
; APPLICANT: Liu, Rihe  
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN  
; FILE REFERENCE: 00786/350007  
; CURRENT APPLICATION NUMBER: US/09/244,796  
; CURRENT FILING DATE: 1998-02-05  
; EARLIER APPLICATION NUMBER: 60/035,963  
; EARLIER FILING DATE: 1997-01-27  
; EARLIER APPLICATION NUMBER: 60/064,491  
; EARLIER FILING DATE: 1997-11-06  
; EARLIER APPLICATION NUMBER: 09/007,005  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17  
; LENGTH: 289  
; TYPE: RNA

Search completed: January 9, 2003, 11:51:00  
Job time : 476 secs

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match          2.9%; Score 31.2; DB 4; Length 289;
Best Local Similarity 9.3%; Pred. No. 12;
Matches 26; Conservative 111; Mismatches 142; Indels 1; Gaps 1;

QY 508 GGAGAAATTCCTTTATCGAGATTCGCGATTACAGAGCGCAATGGGAAGTAACGGTCCTG 567
DB 3 RGRARCRARURURARCRURURURURURARCRURURURURURURURURURURURURUR 62
QY 568 CAACAGAGACAAGTATTGATGGGGAGCTACAAATCCCTGAACGGGATATAGGAAAAATG 627
DB 63 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 122
QY 628 TGTTCCTAAATTCGAAGAATTTAATGACTGCTTGATGCTTAAGAAATCAGAAAGAGACC 687
DB 123 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 182
QY 698 GAAAGCTTCTAAGTAATCAAGAA-AAAATCTTGAGCAATAGCTTTCCATAGAGAAAACC 746
DB 183 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 242
QY 747 GTAGGAAAGGAGTCGATACCAATAAATCCGATTAAAAAA 786
DB 243 RURURURURGRGRGRUAAAAAAAAAAAAAAAAAAAAAA 282

RESULT 50
US-09-134-001C-1809
; Sequence 1809, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1809
; LENGTH: 999
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1809

Query Match          2.9%; Score 31.2; DB 4; Length 999;
Best Local Similarity 53.2%; Pred. No. 21;
Matches 66; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 307 AACAGTAACCAACGACGCTACATAGCATCATGCTTTGTTAGGGTTGTATCGG 366
DB 373 AAAGTGAACAATAGAAATAATTTCTGTTTAATACAGAAATATGTATGTATGAC 432
QY 367 TTGTTTTTTTACTAGGAAGAGCTTGACTTTTCTATGACTTTAGTGAAGAGCTTCCCAT 426
DB 433 ATGTTTTTTTACAGTAAGCGGATTTTATTCTACTGATTTTAGAGGGTATTCTACAAA 492
QY 427 GTTT 430
DB 493 CCTT 496
```

Fri Jan 10 12:01:29 2003

us-09-438-185a-l\_1199590\_1200675.rst

Page 1

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 06:35:28 ; Search time 1940 Seconds  
(without alignments)  
9066.133 Million cell updates/sec

Title: US-09-438-185A-1\_COPY\_1199590\_1200675  
Perfect score: 1086  
Sequence: 1 ttggcaagatcctcaaac.....gtctctgcagtagtgac 1086

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 150 summaries

Database :

EST.\*

- 1: em\_estba.\*
- 2: em\_esthum.\*
- 3: em\_estin.\*
- 4: em\_estnu.\*
- 5: em\_estov.\*
- 6: em\_estpl.\*
- 7: em\_estro.\*
- 8: em\_hic.\*
- 9: gb\_est1.\*
- 10: gb\_est2.\*
- 11: gb\_hic.\*
- 12: gb\_est3.\*
- 13: gb\_est4.\*
- 14: gb\_est5.\*
- 15: em\_estfun.\*
- 16: em\_estom.\*
- 17: gb\_gss.\*
- 18: em\_gss\_hum.\*
- 19: em\_gss\_inv.\*
- 20: em\_gss\_pin.\*
- 21: em\_gss\_vit.\*
- 22: em\_gss\_fun.\*
- 23: em\_gss\_mam.\*
- 24: em\_gss\_mus.\*
- 25: em\_gss\_other.\*
- 26: em\_gss\_pro.\*
- 27: em\_gss\_rod.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query %	Length	ID	Description
C 1	61	5.6	1101	17	CNS0039G	AL063921 Drosophil
C 2	51.4	4.7	997	17	CNS005TE	AL060767 Drosophil
C 3	51.2	4.7	1101	17	CNS0182P	AL108811 Drosophil
C 4	50.2	4.6	1101	17	CNS017KX	AL108171 Drosophil
5	50	4.6	987	17	CNS017LB	AL108185 Drosophil
6	48.4	4.5	1101	17	CNS0106X	AL098595 Drosophil

C	7	48.2	4.4	1001	17	CNS0155H	AL105023 Drosophil
	8	48	4.4	1204	17	CNS016E2	AL106628 Drosophil
	9	47.2	4.3	615	13	BJ345426	BJ345426
	10	47.2	4.3	635	13	BJ345414	BJ345414
	11	46.6	4.3	964	17	CNS006N9	AL065781 Drosophil
	12	45.6	4.2	588	13	BJ135240	BJ135240
	13	45.2	4.2	1101	17	CNS0100X	AL098379 Drosophil
	14	45	4.2	348	13	BJ340360	BJ340360
	15	45	4.1	485	13	BJ340295	BJ340295
C	16	45	4.1	529	9	AU038515	AU038515
	17	45	4.1	546	13	BJ400438	BJ400438
	18	45	4.1	547	13	BJ402020	BJ402020
	19	45	4.1	562	13	BJ376869	BJ376869
	20	45	4.1	584	13	BJ341943	BJ341943
	21	45	4.1	590	13	BJ398603	BJ398603
	22	45	4.1	596	13	BJ376868	BJ376868
	23	45	4.1	597	13	BJ401886	BJ401886
	24	45	4.1	598	13	BJ402538	BJ402538
C	25	45	4.1	653	14	C93147	C93147
	26	45	4.1	654	13	BJ343976	BJ343976
	27	45	4.1	663	13	BJ375941	BJ375941
	28	45	4.1	664	13	BJ401060	BJ401060
	29	45	4.1	665	13	BJ434901	BJ434901
	30	45	4.1	675	13	BJ401293	BJ401293
	31	45	4.1	680	13	BJ343093	BJ343093
	32	45	4.1	688	13	BJ399159	BJ399159
	33	45	4.1	691	13	BJ341290	BJ341290
	34	45	4.1	698	13	BJ401883	BJ401883
	35	45	4.1	718	13	BJ400818	BJ400818
	36	45	4.1	722	13	BJ401195	BJ401195
	37	45	4.1	724	13	BJ341261	BJ341261
	38	45	4.1	734	13	BJ343890	BJ343890
	39	45	4.1	742	13	BJ399243	BJ399243
	40	45	4.1	744	13	BJ347280	BJ347280
	41	45	4.1	760	13	BJ401342	BJ401342
	42	45	4.1	765	13	BJ371948	BJ371948
	43	44.6	4.1	988	17	CNS0072R	AL066743 Drosophil
C	44	44.2	4.1	958	17	CNS0074D	AL066801 Drosophil
	45	43.8	4.0	896	17	CNS016L1	AL106896 Drosophil
	46	43.8	4.0	899	17	AZ550349	AZ550349
C	47	43.8	4.0	914	17	AZ527921	AZ527921
	48	43.8	4.0	914	17	AZ687856	AZ687856
	49	43.8	4.0	924	17	AZ690917	AZ690917
	50	43.8	4.0	944	17	AZ549336	AZ549336
C	51	43.8	4.0	1101	17	CNS014MG	AL104338 Drosophil
	52	43.4	4.0	484	13	BJ344899	BJ344899
	53	43.4	4.0	729	14	C93903	C93903
C	54	43.4	4.0	1101	17	CNS00FMC	AL070972 Drosophil
	55	43.2	4.0	883	13	BM416099	BM416099
C	56	43.2	4.0	1101	17	CNS00B3U	AL056389 Drosophil
	57	42.8	3.9	901	17	CNS00KML	AL078168 Drosophil
C	58	42.8	3.9	1101	17	CNS00ESI	AL069797 Drosophil
	59	42.6	3.9	791	9	AL523778	AL523778
	60	42.4	3.9	859	17	CNS004YY	AL055406 Drosophil
C	61	42.2	3.9	952	17	CNS006V8	AL065967 Drosophil
	62	42.2	3.9	1067	17	CNS00C08	AL058564 Drosophil
C	63	42.2	3.9	1101	17	CNS016T6	AL107172 Drosophil
	64	42	3.9	757	9	AL566489	AL566489
	65	42	3.9	877	9	AL514453	AL514453
	66	42	3.9	1101	17	CNS00FVE	AL071298 Drosophil
C	67	41.8	3.8	886	17	CNS0075DF	AL429961 clone BA0
	68	41.6	3.8	1001	17	CNS0064G	AL062781 Drosophil
C	69	41.6	3.8	1025	17	CNS014J2	AL104216 Drosophil
	70	41.6	3.8	1101	17	CNS0181N	AL108773 Drosophil
	71	41.4	3.8	567	12	BE848353	BE848353
C	72	41.4	3.8	628	10	BB631807	BB631807
	73	41.4	3.8	1101	17	CNS00EEF1	AL069127 Drosophil
C	74	41.2	3.8	801	17	AQ050553	AQ050553
	75	41.2	3.8	905	17	CNS00D59	AL060243
C	76	41.2	3.8	1067	13	BJ600924	BJ600924
	77	41	3.8	748	13	BJ141956	BJ141956
C	78	41	3.8	757	9	AL514421	AL514421
	79	41	3.8	772	17	BH596443	BH596443





Db 797 ADDRDAADDDAAATTTTWTWTTTDDDKWKTDTWTEWAADRTWDRDDDDDRDRAGT 738  
QY 713 AATCTTGGAACAATAGCTTTTCATAGAGAAACCGTCAGAGAGAGTCGATACCAATAAA 772  
Db 737 AGRKWRRTWKRRKRDRTKWDADDDTAKDRRRRGGDGDAGAGKKTGKRKRDRDRAT 678  
QY 773 TTCGGATTAATAAAGGAGTGTTCACACACCCATATCATCTGAGAGCAACCTTCCCTTTGA 832  
Db 677 WDRTDAMWADAAMWTTTDDTDWDRKRRRGARRRRRTTAAAWDWMTWRAWDKWD 618  
QY 833 TTCCTTTGATATGACAGAGAATGACAAAGAGTAAGTTTTTGGAAAAATCTTAAAAATGAA 892  
Db 617 WKTRADRWRDWAADTTWDKADRDWAKARARARRDRARAARDRRTTKGKTTTATWT 558  
QY 893 ATGCATTTGTCTGAGCTAAAGCTTCTCTCTCTTTATTTTCTCTCTCTCTCTCTCTCGG 952  
Db 557 TWAARAAMWAWWATTTATTTTWTWTTTWTWTTTWTWTTTWTWTTTWTWTTTAAWAAWTTAAW 498  
QY 953 TAGTAGGAATA 963  
Db 497 TAAWAAAAA 487  
RESULT 2  
CNS0005TE/c 997 bp DNA linear GSS 03-JUN-1999  
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #  
DEFINITION BAC12K22 of RPI1-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION AL060767.1 GI:4943573  
VERSION 1  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster.  
REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 997)  
Direct Submission  
AUTHORS Aaron Mammosier in Pieter de Jong's laboratory in the Department of  
TITLE Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
JOURNAL N.Y. The library is named RPI1-98 and was constructed by partial  
restriction digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain Y2, cn bw SP, the same strain used for the BDGP's  
P1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones the entire library or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).  
Location/Qualifiers  
FEATURES  
source 1..997  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="BAC12K22"  
/clone\_lib="RPI1-98"  
/note="end : TET3"  
BASE COUNT 89 a 99 c 13 g 258 t 538 others  
ORIGIN  
Query Match 4.7%; Score 51.4; DB 17; Length 997;  
Best Local Similarity 18.6%; Pred. No. 0.019;  
Matches 103; Conservative 162; Mismatches 288; Indels 1; Gaps 1;

QY 409 AGTGAAGAGCTTCCCATGTTTATGAAAAATTCAGAAAAATGAGGATGTTAGAACCCAAAG 468  
Db 918 RRAGRAGRGRSAGGGRGRRRAAARADARARARARARARARARARARARARARARAR 859  
QY 469 CAGCTGCCCTAAAGGTCATGGATCAGATCAGAGTGAAGGAGAAATTTATCTTATCGAG 528  
Db 858 GRRRRGRAGRRRAGGAGGRAGRRGRGRRRRARARRRAGARARARARARARARARAR 800  
QY 529 AGTTCGCATTACAGAGCAATCGGAAAGTAAACGCTCTCTGCAACAGAGACAACTATTCGTG 588  
Db 799 RRRAGRRRRRRCAGRRRRRRGRRRRRRRRAARRARARARARARARARARARAR 740  
QY 589 GGGAGCTCAAAATCTGAAACGGGATATAGGAAAAATGTGTTCTTAATCGAAGATTT 648  
Db 739 RAGAAFRARRRRRRRRRGAGARARARARRRRAARAGAGARRRRRRRRGRRGARRRG 680  
QY 649 AATGACTGTTGATGGTCTAAGAAATCAGAAAGAGAGCCCAAGCTTCTAAGTAATCAAG 708  
Db 679 RGRRRRAGARRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 620  
QY 709 AAAAATCTTTGGACAATAGCTTTTCCATAGAGAAACCGTCGAGAAAGAGTCGATACCA 768  
Db 619 ARGRRRAGRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 560  
QY 769 TAAATTCGGATTAAAAAGGGAGTCTTTACAAACAGCCATCATCGTGGAGCAACCTTCCCT 828  
Db 559 RRRARRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 500  
QY 829 TTGATCTTTTATAATCAGACAGATGAACAAAGAGTAAGTTTTTGAAGAAATCTTAAAT 888  
Db 499 GRRRAGARRRRRAAARARARARAGARRRRRRRRRRRRRRRRRRRRRRRRRRRR 440  
QY 889 AGAATGCATTGTCTGAGCTTAAAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 948  
Db 439 NNNNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 380  
QY 949 TCGGTAGTAGAAT 962  
Db 379 TTTTNTNTNTNTNT 366  
RESULT 3  
CNS0182P 1101 bp DNA linear GSS 26-JUL-1999  
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC  
DEFINITION BAC37D10 of DrosBAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION AL108811  
VERSION AL108811.1 GI:5629115  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster.  
REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1101)  
Direct Submission  
AUTHORS Submitted (23-JULY-1999) Genoscope - Centre National de Sequencage :  
TITLE BP 191 91006 EVRY cedex - FRANCE (E-mail : [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr))  
JOURNAL - Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
COMMENT Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
<http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelosBAC11.  
Location/Qualifiers  
FEATURES  
source 1..1101  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"



http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.

```
FEATURES             Location/Qualifiers
    source             1..987
                        /organism="Drosophila melanogaster"
                        /db_xref="taxon:7227"
                        /clone="BACN37A07"
                        /clone_lib="DrosBAC"
                        /plasmid="pBelOBAC11"
                        /notes="end : SP6"
BASE COUNT           224 a 122 c 180 g 211 t 250 others
ORIGIN
Query Match          4.8%; Score 50; DB 17; Length 987;
Best Local Similarity 30.5%; Pred. No. 0.044; Indels 0; Gaps 0;
Matches 96; Conservative 77; Mismatches 142;
QY 626 TGTGTTCTTAATCGAAGATTTAATGACTGCTGATGCTGCTAAGAAATCAGAAAGGAGA 685
Db 634 TGAGTCTCKGKWKGTWTTTWWGKGGKGGTGGGDKGKTGWTGTYTDCVAGGAGG 693
QY 686 CGGAAGCTTCTAAGTAATCAAGAAAAATCTTGGCAATAGCTTTTCCATAGAGAAAC 745
Db 694 RAGWAARGADSGKMWAAANGRWAAAAARTMTTMMWMTHTHHHARDAAWAAAAWG 753
QY 746 CGTGGAGGAAGAGTTCGATACCAATAATCCGATTAAAAAGGAGGCTGTACACAGCC 805
Db 754 KGTGTAARAAATDBYTWGGGGKKGDTGKWKWADAHWDTKATAAWKAAATAATW 813
QY 806 ATATCAGTGGCAACCTTCCCTTTGATTCTTTGATATGATGACAGAGATGAACAAAGAT 865
Db 814 AAAAATATKMTWTTAAATWTKATWRTGTTTAAWAWDWTGDKGGGGDRRAAAG 873
QY 866 AAGTCTTTCAAAAATCTCAAAATAGAAATGCAATTCGTGCGAGCTAAAGCTTCCTCTCT 925
Db 874 WRATATKTDTTTBTYYTMTWYVYTCHTTTTTTTTTTTTTTTTTTTTTTCTTCTTCT 933
QY 926 TTATTTCTCTTTGT 940
Db 934 YTTTTTYYTYYT 948
```

```
RESULT 6
CNS0106X             1101 bp DNA linear GSS 26-JUL-1999
LOCUS                Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION            BACN03K20 of DrosBAC library from Drosophila melanogaster (fruit
                        fly), genomic survey sequence.
ACCESSION             AL098595
VERSION               AL098595.1 GI:5610206
KEYWORDS              GSS.
SOURCE                Drosophila melanogaster.
ORGANISM              Drosophila melanogaster.
                        Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
                        Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
                        Ephydroidea; Drosophilidae; Drosophila.
REFERENCE              1 (bases 1 to 1101)
AUTHORS               Direct Submission
TITLE                 Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
JOURNAL               BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
                        - Web : www.genoscope.cns.fr)
COMMENT               Determination of this BAC-end sequence was carried out as part of a
                        collaboration with the European Drosophila Genome Project (EDGP) -
                        http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
                        library (Dros BAC) was made by Alain Billaud at CEPH (Centre
                        d'Etude du Polymorphisme Humain) with funding provided by a MRC
                        project grant. The DNA was prepared from embryos by Alain Bucheton
                        and Genevieve Payan. It has been constructed in the vector
```

```
FEATURES             Location/Qualifiers
    source             1..1101
                        /organism="Drosophila melanogaster"
                        /db_xref="taxon:7227"
                        /clone="BACN03K20"
                        /clone_lib="DrosBAC"
                        /plasmid="pBelOBAC11"
                        /notes="end : T7"
BASE COUNT           258 a 107 c 60 g 175 t 501 others
ORIGIN
Query Match          4.8%; Score 48.4; DB 17; Length 1101;
Best Local Similarity 19.4%; Pred. No. 0.11;
Matches 83; Conservative 178; Mismatches 163; Indels 4; Gaps 1;
QY 480 AAAGTCATGATGAGATCAGGAGTAAAGGAGAAATATCTTTATCGAGAGTTCGCATTA 539
Db 661 ARKDKDTAKDWKDKATKAKAKAKAKAKAKAKATATGDKAKAKAKAKAKAKAKAKAKAK 716
QY 540 CAGAGCAATGGGAAGTAAGCGTCTGCAACAGACAGCAAGTATTGATGGGAGCTACAA 599
Db 717 RAKARAWDTATWTATAADAAADGAKDKAKAKAKADAROKRRWDKAKRRRAKAKAKA 776
QY 600 ATCTCTGAACGGGATAATAGGAAAAATGTGTTCTTAATTCGGAAGAAATTTAATGACTCT 659
Db 777 ADDADADAKADAKADAKADADADGCGKDKKRAKDKKDKKDKKDKKDKKDKKDKKDK 836
QY 660 GATGCTCTAAGAAATCAGAAAGGAGCGCAAGCTTCTAAGTAATCAAGAAAAATCTTG 719
Db 837 AAAKAKADAAAGADAKAKRRRAGDKAKADAKAKAKAKAKAKAKAKAKAKAKAKAKAK 896
QY 720 GACAATAGCTTTTCCATAGAGAAACCGTGGAGGAAGAGTTCGATACCAATAATTCGAT 779
Db 897 AKAKKKKKKKDKAKAKAKAKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDK 956
QY 780 TAAAGAGGAGTGTCTACACAGCATATCAGTGGAGCAACCTTCCTCTTGTGATCTTTG 839
Db 957 KDKAKDKDKDKDKADAKAWAKADADAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAK 1016
QY 840 ATAATGACAGAGTGTGACAAAGAGTAAAGTGTGTTTGAAGAAATTTCTAAATAGAAATGCAT 899
Db 1017 AKADDAKAKAKADAKAADAADAKAAKAKAKAKADADADAKAKAKAKAKAKAKAKAK 1076
QY 900 TGTGTGCA 907
Db 1077 AKKAKAKD 1084
```

```
RESULT 7
CNS0155H             1001 bp DNA linear GSS 26-JUL-1999
LOCUS                Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION            BACN13C23 of DrosBAC library from Drosophila melanogaster (fruit
                        fly), genomic survey sequence.
ACCESSION             AL105023
VERSION               AL105023.1 GI:5617037
KEYWORDS              GSS.
SOURCE                Drosophila melanogaster.
ORGANISM              Drosophila melanogaster.
                        Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
                        Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
                        Ephydroidea; Drosophilidae; Drosophila.
REFERENCE              1 (bases 1 to 1001)
AUTHORS               Direct Submission
TITLE                 Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
JOURNAL               BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
                        - Web : www.genoscope.cns.fr)
COMMENT               Determination of this BAC-end sequence was carried out as part of a
                        collaboration with the European Drosophila Genome Project (EDGP) -
                        http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
                        library (Dros BAC) was made by Alain Billaud at CEPH (Centre
```

d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.

FEATURES  
source  
Location/Qualifiers  
1..1001  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="BACN13C23"  
/clone\_lib="DrosBAC"  
/plasmid="pBelOBAC11"  
/note="end : SP6"  
BASE COUNT 266 a 219 c 134 g 150 t 232 others  
ORIGIN  
Query Match 4.4%; Score 48.2; DB 17; Length 1001;  
Best Local Similarity 28.5%; Pred. No. 0.13;  
Matches 145; Conservative 106; Mismatches 257; Indels 0; Gaps 0;  
Qy 380 GGAAGAGCTTGACTTTTCTATGACTTTTATGAGAGTCTTCCCATCTTTATGAAAAAT 439  
Db 494 SGRGCGGGGTDTYKKBKBBBYKKGKGMGADARAARKTCACTAAAWRAAGACAAA 553  
Qy 440 TCAGAAATGAGGGATGTAGAGCAAGCGCGCTTAAAGGTCTATGATGATGATCA 499  
Db 554 AGRRVRARWWDAGAGAAGAGATNAGAAAAAAGRGMAAABDAGRASAAARKTKWG 613  
Qy 500 GGAGTTAAGGAGAAATTTATCTTTCGAGAGTTCGCATTACAGAGCAATGGGAAGTAA 559  
Db 614 ARGAAAAAGWDADDAATWRRKAARAADADWAAAAACAANAANVHNAWAARAKAA 673  
Qy 560 CGTCTCTCAACAGACAGCAAGTATGATGGGAGCTTCAAACTCTGAAACGGGATATAG 619  
Db 674 ARRAAAAAAGADAAAAAAGRAAARDARAANAATAAAMWAATAARWAARAKDAR 733  
Qy 620 GAAAAATGTTTCTTAATCGAAGAAATTTATGACTGCTTGATGCTCTAAGAAATCAGAA 679  
Db 734 AAAAAAATTAATTAATAAAMWAAMWAAMWAAMWAAMWAAMWAAMWAAMWAAMWAAMWA 793  
Qy 680 AGGAGCCGAAAGCTTCTAAGTAATCAAGAAAAAATCTTGGCAATAGCTTTCCATAGA 739  
Db 794 AATAATTTAAWAAAAAATTAATAAATTTAAWAAAAAATTAATAAATTTAAWAAAAAT 853  
Qy 740 GAAACCGTCGAGGAAGGTCGATACCATATAAATTCGGATTAAAAAAGGGAGTCTTACA 799  
Db 854 AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 913  
Qy 800 ACAGCATATCAGTCGAGCAACCTTCCCTTTGATCTTTGATAATGACAGAGATGAACA 859  
Db 914 WAWADTTTWTWTAWAAATAATWTAATTAATTAATAATAATAATAATAATAATAATW 973  
Qy 860 AAGAGTAAGCTTTTCAAAAAATCTTAAA 887  
Db 974 AWAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1001

RESULT 8  
CNS016E2/c 1204 bp DNA linear GSS 26-JUL-1999  
LOCUS  
DEFINITION  
Drosophila melanogaster genome survey sequence T7 end of BAC  
BACN15A12 of DrosBAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION  
AL106628  
VERSION  
AL106628.1 GI:5622852  
KEYWORDS  
GSS.  
SOURCE  
Drosophila melanogaster.  
ORGANISM  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
1. (bases 1 to 1204)  
Genoscope.  
AUTHORS  
TITLE  
Direct Submission

JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelOBAC11.

FEATURES  
source  
Location/Qualifiers  
1..1204  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="BACN15A12"  
/clone\_lib="DrosBAC"  
/plasmid="pBelOBAC11"  
/note="end : T7"  
BASE COUNT 298 a 172 c 106 g 316 t 312 others  
ORIGIN

Query Match 4.4%; Score 48; DB 17; Length 1204;  
Best Local Similarity 18.2%; Pred. No. 0.14;  
Matches 66; Conservative 146; Mismatches 150; Indels 0; Gaps 0;  
Qy 699 AGTAATCAAGAAAAATCTTGGACAAATAGCTTTTCCATAGAGAAACCTTGAGGAAGGA 758  
Db 1189 AKWAKWAKATAAKWAKAKAKWAKAAAAAANAANAANAANAANAANAANAANAANA 1130  
Qy 759 GTCGATACCATTAATTCGGATTAAAAAGGAGTGTTCACACAGCCATATCAGTGGAGC 818  
Db 1129 AKAADDAADAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 1070  
Qy 819 AACCTTCCCTTTGATCTTGTATGATGACAGAGATGAACAAGAGTAAGTTTGTGAAA 878  
Db 1069 DDRDKAAADAARAKKAKADADAAADKDKKAKADAWDDADDAARAKKAKKDKKDD 1010  
Qy 879 ATTCTAAAATAGAAATGCTTGTGTCGAGCTTAAAGTCTTCTTCTTATTTTCTTTT 938  
Db 1009 ATDADWAAAKAKKDKKAWKKKDKKAKKAKKAWKKKDKKDKKDKKDKKDKKDKK 950  
Qy 939 GTAGAATGATTCGGTAGTAGGAATATGGGTGAGTATGAGCAGTGTGAGTGTGCTG 998  
Db 949 AAADAAAAAARAKKDKKDKKDKKDKKDKKDKKDKKDKKDKKDKKDKKDKKDKKDK 890  
Qy 999 TCTGGAGAACAGGAAAGTATGCTCTCAGCTTTAGAGCAATCTTTCAGATATATCTTG 1058  
Db 889 AAKKDKKDKDAADAKKAKDKKDKKDKKDKKDKKDKKDKKDKKDKKDKKDKKDKK 830  
Qy 1059 GG 1060  
Db 829 GG 828

RESULT 9  
BJ345426 615 bp mRNA linear EST 06-MAR-2002  
LOCUS  
DEFINITION  
BJ345426 Dictyostellium discoideum cDNA library, AF Dictyostellium  
discoideum cDNA clone dda23117 3', mRNA sequence.  
ACCESSION  
BJ345426  
VERSION  
BJ345426.1 GI:19215933  
KEYWORDS  
EST.  
SOURCE  
Dictyostellium discoideum.  
ORGANISM  
Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.  
REFERENCE  
1. (bases 1 to 615)  
AUTHORS  
Urushihara, H., Tanaka, Y., Kohara, Y., and Shin-i, T.  
TITLE  
Full length cDNA of Dictyostellium discoideum at the aggregation  
stage  
JOURNAL  
Unpublished (2002)  
CONTACT: Tadasu Shin-i  
COMMENT  
Center For Genetic Resource Information

us-09-438-185a-1\_1199590\_1200675.rst

'Fri Jan 10 12:01:29 2003

National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.  
Location/Qualifiers

# FEATURES

Source  
1. .615  
/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone="dda23117"  
/clone\_lib="Dictyostelium discoideum cDNA library, AF"  
/sex="mat A"  
/dev\_stage="Aggregation stage"  
207 a 94 c 103 g 210 t 1 others

## BASE COUNT

ORIGIN  
Query Match 4.3%; Score 47.2; DB 13; Length 615;  
Best Local Similarity 59.8%; Pred. No. 0.22;  
Matches 79; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 193 TAAAGGAGACACGTAGCTATCAATGAAGCGTGTCCAAAGTTCCTGAGGAGAACTGAT 252

Db 220 TAAATGGTAATTTTGTGTTATACAGTTAAAGAGTTCACCTGTTGATGATAA 279

Qy 253 AAGAAGCGTCCATATGCTTCTTCCTTCATCAATAGTCGCTTCAACAGT 312

Db 280 AATACCTGCACCAATGCTCTTAATGTATGACCTCTTTACATATCCAAATCAACTGT 339

Qy 313 AAACCAAAAGCA 324

Db 340 AAACCAATAACA 351

## RESULT 10

BJ345414 635 bp mRNA linear EST 06-MAR-2002

LOCUS BJ345414 Dictyostelium discoideum cDNA library, AF Dictyostelium

DEFINITION dictyostelium cDNA clone dda2314 3', mRNA sequence.

ACCESSION BJ345414

VERSION BJ345414.1 GI:19215921

KEYWORDS EST.

SOURCE Dictyostelium discoideum.

ORGANISM Dictyostelium discoideum.

REFERENCE Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.

Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.

Full length cDNA of Dictyostelium discoideum at the aggregation

stage

Unpublished (2002)

Contact: Tadasu Shin-i

Center for Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

Location/Qualifiers

1. .635

/organism="Dictyostelium discoideum"

/strain="AX4"

/db\_xref="taxon:44689"

/clone="dda2314"

/clone\_lib="Dictyostelium discoideum cDNA library, AF"

/sex="mat A"

/dev\_stage="Aggregation stage"

216 a 95 c 105 g 218 t 1 others

## BASE COUNT

ORIGIN

Query Match 4.3%; Score 47.2; DB 13; Length 635;

Best Local Similarity 59.8%; Pred. No. 0.22;

Matches 79; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 193 TAAAGGAGACACGTAGCTATCAATGAAGCGTGTCCAAAGTTCCTGAGGAGAACTGAT 252

Db 218 TAAATGGTAATTTTGTGTTATACAGTTAAAGAGTTCACCTGTTGATGATAA 277

Qy 253 AAGAAGCGTCCATATGCTTCTTCCTTCATCAATAGTCGCTTCAACAGT 312

Db 278 AATACCTGCACCAATGCTCTTAATGTATGACCTCTTTACATATCCAAATCAACTGT 337

Qy 313 AAACCAAAAGCA 324

Db 338 AAACCAATAACA 349

## RESULT 11

CNS006N9 964 bp DNA linear GSS 03-JUN-1999

LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #

DEFINITION BACR14J19 of RPCI-98 library from Drosophila melanogaster (fruit

fly) genomic survey sequence.

ACCESSION AL065781

VERSION AL065781.1 GI:4944661

KEYWORDS GSS.

SOURCE Drosophila melanogaster.

ORGANISM Drosophila melanogaster.

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 964)

Genoscope.

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Aaron Mammeter in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCI-98 and was constructed by partial

ECORI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain y2; cn bw sp, the same strain used for the BDGP's

pl and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

Location/Qualifiers

1. .964

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="BACR14J19"

/clone\_lib="RPCI-98"

/notes="end : T7"

BASE COUNT 409 a 111 c 71 g 130 t 243 others

## ORIGIN

Query Match 4.3%; Score 46.6; DB 17; Length 964;

Best Local Similarity 35.2%; Pred. No. 0.32;

Matches 115; Conservative 55; Mismatches 157; Indels 0; Gaps 0;

Qy 581 TATTGATGGGAGCTACAAATCCTGAACGGGATATAGGAAAATGTCTTCTAATTCG 640

Db 457 TATVMMMTATATMKATAAAAAKAAATATATGTTAWGAAAAAADAGTKAATAADAA 516

Qy 641 AAGAAATTAATGACTCCTTGTATGGTCTTAAGAAATCAGAAGGAGCCGGAAGCTTCAAG 700

Db 517 AAAAAAANAAAAAANKAAWKKTAATAAAAAAATAAAAAAATAAAAKTTATATR 576

Qy 701 TAATCAAGAAAAAATCTTGGCAATAGCTTTTCCATAGAGAAAAACCGTCAGGAAGAGT 760

Db 577 TAAAAAANAAAAAADDGGTGTATATTTGTTAKRTAWDAAAAAATTTATAAAAAA 636

```

Qy 761 CGATACATAAATCCGATTAAAAAGGGAGTCTTACAACGCCATCATCAGTGAGCAA 820
    ||| | ||| : : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 637 AAATAAAAAAANDKAKWKAATATATAAAATGTAGTAWTAKTAAAAATATAAAAA 696
    : : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Qy 821 CCTTCCCTGTTGTTCTTGTATGATGACAGAGATGAACAAAGAGTAAAGTTTGTGAAAAAT 880
    : : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 697 ADAADAADAADAAKTKWKAADAADAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 756
    : : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Qy 881 TCTAAAAATGAATCATTTGTTGTCGA 907
    : : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Db 757 DAAAAADTAKAKWDNADAKDDDKRAGA 783

RESULT 12
LOCUS BJI135240 588 bp mRNA linear EST 23-JAN-2002
DEFINITION BJI135240 unpublished oligo-capped cDNA library, C. elegans L1 stage
Caenorhabditis elegans cDNA clone yk1097h08 3', mRNA sequence.
ACCESSION BJI135240.1 GI:18295397
VERSION EST.
KEYWORDS Caenorhabditis elegans.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1 (bases 1 to 588)
AUTHORS Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE A complementary view of the C.elegans genome
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@gene.nig.ac.jp.
FEATURES
    source
    1..588
    /organism="Caenorhabditis elegans"
    /strain="N2"
    /db_xref="taxon:6239"
    /clone="yk1097h08"
    /clone_lib="unpublished oligo-capped cDNA library, C.
    elegans L1 stage"
    /sex="hermaphrodite"
    /tissue_type="whole animal"
    /dev_stage="L1"
    /note="The AD-wrmcDNA library was generated with poly(A)+
    RNA isolated from both hermaphrodite and male N2 worms of
    all larval stages, embryos, adults and dauers and the
    subsequent generation of cDNAs by poly(A) priming. The
    cDNAs were cloned into pPC86"
BASE COUNT 166 a 112 c 131 g 179 t
ORIGIN

Query Match 4.2%; Score 45.6; DB 13; Length 588;
Best Local Similarity 59.1%; Pred. No. 0.56;
Matches 78; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy 191 TCTAAGGAGAACACGACGTATATCAATGAAAGCGGTGTCGAAGTTCCTGAGGAACTG 250
    ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 261 TCAAAATCTACAACTCTCCGGCTTACTACTCAACGCATATTTGAAGCTCTCCAAATGAAC 320
    : : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Qy 251 ATAAGACGGCTCCATATGCTTTCTTCTTCCTTCATGTTTTCATTAAGTCGCGCTTCAACA 310
    ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 321 AAAAGTCGCGCTCCGTAAGCTTTTCTTCCCATCTTGTGACAGATTCCAAATTCGATT 380
    : : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
Qy 311 GTAACCAAAAG 322
    ||| ||| ||| |||
Db 381 GTCAACCAAGTAG 392

```

```

RESULT 13
LOCUS CNS0100X 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN03G04 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL098379.1 GI:5609990
VERSION GSS.
KEYWORDS Drosophila melanogaster.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
JOURNAL Genoscope.
COMMENT Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.
FEATURES
    Location/Qualifiers
    1..1101
    /organism="Drosophila melanogaster"
    /db_xref="taxon:7227"
    /clone="BACN03G04"
    /clone_lib="DrosBAC"
    /plasmid="pBelobAC11"
    /note="end : SP6"
BASE COUNT 195 a 108 c 131 g 161 t 506 others
ORIGIN

Query Match 4.2%; Score 45.2; DB 17; Length 1101;
Best Local Similarity 17.6%; Pred. No. 0.74;
Matches 94; Conservative 212; Mismatches 226; Indels 1; Gaps 1;

Qy 357 TTGTATGCGCTGTTTTTACTAGGAAGAGCTTGTACTTTTCTATGACTTTAGTGAAGA 416
    : : | | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Db 463 WTWTTTTGKKWDTKTWKDDGGTTWTDTWKDKTKDXTKTWTDVWADNAGWWTOKD 522
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Qy 417 GTCTTCCCATGTTTATGAAAATTCAGAAATGAGGATGTAGAAGCCAGGCGTGCC 476
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Db 523 KTRAKKRTKTDTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK 582
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Qy 477 CTAAAGGTCATGGATGAGATCAGGAGTTAAGGAGAAATATCTTTATCGAGAGTTCCGA 536
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Db 583 DADGKADTATKTDGWTGTTTADTTDDADWDKWDKADTKAGRRGWWGKGG 642
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Qy 537 TTACAGAGCAATGGGAAAGTAACGGTCCTGCAACAGACAGACAAGTATTCATGGGAGCTA 596
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Db 643 WKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK 702
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Qy 597 CAAATCCTGAAAGCGGATAATAGGAAAAATGTCTTCTTAATTCGAAGAAATTAATGACTG 656
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Db 703 AAWRAKATRAAADAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 762
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Qy 657 CTGTAGTGGTCTAAGAAATCAGAAAGGAGAGCCGAAAGCTTCTAAGTAAATCAAGAAAAAT 716
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Db 763 ARKADRDDDDKAKRDGDKGDWDADKDKRAGRDODDTTKW-GWKDRAGKKKKAKKDH 821
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Qy 717 TTGCAACAATAGCTTTTCCATAGAGAAAAACCGTCGAGGAAGAGCTCGATACCAATAATTC 776
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Db 822 KTAWWNADDTGTAWATWAKTATAKTKAKATKDKBKRGRRGGGGRGRRGRRKRGRKR 881
    : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Qy 777 GATTAATAAAGGAGGTGCTTACAACAGCCATATCATGTTGGAGCAACCTTCCCTTGATTCT 836

```



Full length cDNA of Dictyostereum discoideum at the slug stage  
unpublished (2002)  
Contact: Tadasu Shin-i

[illegible][illegible]



us-09-438-185a-l\_1199590\_1200675.rst

Fri Jan 10 12:01:29 2003

```

Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1..590
/organism="Dictyostellium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddsin01"
/clone_lib="Dictyostellium discoideum cDNA library, SF"
/sex="mat A"
/dev_stage="Slug stage"
BASE COUNT 199 a 89 c 97 g 205 t
ORIGIN
Query Match 4.1%; Score 45; DB 13; Length 590;
Best Local Similarity 61.5%; Pred. No. 0.8;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 208 TACGTTATCAATGAAGCGTGTCCCAAGTTCTCTGAGGAGAACTGATAAGAACGGCTCCATA 267
Db 231 TGCCTTATCAGTTAAAGAGTGTCCATTTCACCTGTGTGATGATAAAATACCTGCACCATA 290
QY 268 TGCCTTTCTTCCTTCATGTTTCAATTAAGTCCGCTTTCAACAGTAAACCAAGCA 324
Db 291 TGCCTAATTGTATCCTCTTCTTACATAATCCAAATTCACCTGATAACCAATAACA 347

RESULT 22
BJ376868 596 bp mRNA linear EST 08-MAR-2002
LOCUS BJ376868 Dictyostellium discoideum cDNA library, CF Dictyostellium
DEFINITION BJ376868 Dictyostellium cDNA clone ddc30g09 3', mRNA sequence.
ACCESSION BJ376868
VERSION BJ376868.1 GI:19286251
KEYWORDS EST.
SOURCE Dictyostellium discoideum.
ORGANISM Dictyostellium discoideum.
REFERENCE Dictyostellium discoideum.
AUTHORS Eukaryota: Mycetozoa; Dictyostellida; Dictyostellium.
TITLE Urushihara, H., Tanaka, Y., Kohara, Y., and Shin-i, T.
Full length cDNA of Dictyostellium discoideum at the culmination
stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1..584
/organism="Dictyostellium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dda8h19"
/clone_lib="Dictyostellium discoideum cDNA library, AF"
/sex="mat A"
/dev_stage="Aggregation stage"
BASE COUNT 199 a 91 c 101 g 192 t 1 others
ORIGIN
Query Match 4.1%; Score 45; DB 13; Length 584;
Best Local Similarity 61.5%; Pred. No. 0.8;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 208 TACGTTATCAATGAAGCGTGTCCCAAGTTCTCTGAGGAGAACTGATAAGAACGGCTCCATA 267
Db 178 TGCCTTATCAGTTAAAGAGTGTCCATTTCACCTGTGTGATGATAAAATACCTGCACCATA 237
QY 268 TGCCTTTCTTCCTTCATGTTTCAATTAAGTCCGCTTTCAACAGTAAACCAAGCA 324
Db 238 TGCCTAATTGTATCCTCTTCTTACATAATCCAAATTCACCTGATAACCAATAACA 294

RESULT 21
BJ398603 590 bp mRNA linear EST 10-MAR-2002
LOCUS BJ398603 Dictyostellium discoideum cDNA library, SF Dictyostellium
DEFINITION BJ398603 Dictyostellium cDNA clone ddsin01 3', mRNA sequence.
ACCESSION BJ398603
VERSION BJ398603.1 GI:19311520
KEYWORDS EST.
SOURCE Dictyostellium discoideum.
ORGANISM Dictyostellium discoideum.
REFERENCE Dictyostellium discoideum.
AUTHORS Eukaryota: Mycetozoa; Dictyostellida; Dictyostellium.
TITLE Urushihara, H., Tanaka, Y., Kohara, Y., and Shin-i, T.
Full length cDNA of Dictyostellium discoideum at the slug stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan

```





```

RESULT 29
BJ434901
LOCUS          665 bp      mRNA      linear      EST 13-MAR-2002
DEFINITION    Dictyostelium discoideum cDNA library, VF Dictyostelium
VERSION       BJ434901
ACCESSION    BJ434901
KEYWORDS     Dictyostelium discoideum.
SOURCE       Dictyostelium discoideum.
ORGANISM     Dictyostelium discoideum.
REFERENCE    1 (bases 1 to 665)
AUTHORS      Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE        Full length cDNA of Dictyostelium discoideum at the vegetative
JOURNAL      stage
COMMENT      Unpublished (2002)
              Contact: Tadasu Shin-i
              Center For Genetic Resource Information
              National Institute of Genetics
              1111 Yata, Mishima, Shizuoka 411-8540, Japan
              Tel: 81-559-81-6856
              Fax: 81-559-81-6855
              Email: tshini@genes.nig.ac.jp.
FEATURES     source
              1..665
              /organism="Dictyostelium discoideum"
              /strain="AX4"
              /db_xref="taxon:44689"
              /clone_lib="Dictyostelium discoideum cDNA library, VF"
              /sex="mat A"
              /dev_stage="Growth phase"
BASE COUNT   222 a 99 c 113 g 230 t 1 others
ORIGIN
Query Match 4.1%; Score 45; DB 13; Length 665;
Best Local Similarity 61.5%; Pred. No. 0.8;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 208 TACGTTATCAATGAAGCGGTGCCAAGTTCCTGAGGAGAACTGATAAGACGGCTCCATA 267
DB 236 TGCCTTATCAGTTAAAGAGTGTTCACCTGTTGATGATAAAATACCTGCACCATA 295
QY 268 TGCCTTTCTTCCTTCATGCTTTTCAATAAGTCCGCTTCAACAGTAACCAAAAGCA 324
DB 296 TGCCTAAATTGATACCTCTTTACATAATCCAAATTCAACTGTAACCAATAACA 352

RESULT 30
BJ401293
LOCUS          675 bp      mRNA      linear      EST 10-MAR-2002
DEFINITION    Dictyostelium discoideum cDNA library, SF Dictyostelium
VERSION       BJ401293
ACCESSION    BJ401293
KEYWORDS     Dictyostelium discoideum.
SOURCE       Dictyostelium discoideum.
ORGANISM     Dictyostelium discoideum.
REFERENCE    1 (bases 1 to 675)
AUTHORS      Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE        Full length cDNA of Dictyostelium discoideum at the slug stage
JOURNAL      Unpublished (2002)
COMMENT      Contact: Tadasu Shin-i
              Center For Genetic Resource Information
              National Institute of Genetics
              1111 Yata, Mishima, Shizuoka 411-8540, Japan
              Tel: 81-559-81-6856
              Fax: 81-559-81-6855
              Email: tshini@genes.nig.ac.jp.
FEATURES     source
              1..675
              /organism="Dictyostelium discoideum"
              /strain="AX4"
              /db_xref="taxon:44689"
              /clone_lib="Dictyostelium discoideum cDNA library, SF"
              /sex="mat A"
              /dev_stage="Aggregation stage"
BASE COUNT   225 a 99 c 122 g 232 t 2 others
ORIGIN
Query Match 4.1%; Score 45; DB 13; Length 680;
Best Local Similarity 61.5%; Pred. No. 0.81;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 208 TACGTTATCAATGAAGCGGTGCCAAGTTCCTGAGGAGAACTGATAAGACGGCTCCATA 267
DB 185 TGCCTTATCAGTTAAAGAGTGTTCACCTGTTGATGATAAAATACCTGCACCATA 244
QY 268 TGCCTTTCTTCCTTCATGCTTTTCAATAAGTCCGCTTCAACAGTAACCAAAAGCA 324
DB 245 TGCCTAAATTGATACCTCTTTACATAATCCAAATTCAACTGTAACCAATAACA 301

RESULT 31
BJ343093
LOCUS          680 bp      mRNA      linear      EST 07-MAR-2002
DEFINITION    Dictyostelium discoideum cDNA library, AF Dictyostelium
VERSION       BJ343093
ACCESSION    BJ343093
KEYWORDS     Dictyostelium discoideum.
SOURCE       Dictyostelium discoideum.
ORGANISM     Dictyostelium discoideum.
REFERENCE    1 (bases 1 to 680)
AUTHORS      Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE        Full length cDNA of Dictyostelium discoideum at the aggregation
JOURNAL      stage
COMMENT      Unpublished (2002)
              Contact: Tadasu Shin-i
              Center For Genetic Resource Information
              National Institute of Genetics
              1111 Yata, Mishima, Shizuoka 411-8540, Japan
              Tel: 81-559-81-6856
              Fax: 81-559-81-6855
              Email: tshini@genes.nig.ac.jp.
FEATURES     source
              1..680
              /organism="Dictyostelium discoideum"
              /strain="AX4"
              /db_xref="taxon:44689"
              /clone_lib="Dictyostelium discoideum cDNA library, AF"
              /sex="mat A"
              /dev_stage="Aggregation stage"
BASE COUNT   225 a 99 c 122 g 232 t 2 others
ORIGIN
Query Match 4.1%; Score 45; DB 13; Length 680;
Best Local Similarity 61.5%; Pred. No. 0.81;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 208 TACGTTATCAATGAAGCGGTGCCAAGTTCCTGAGGAGAACTGATAAGACGGCTCCATA 267
DB 185 TGCCTTATCAGTTAAAGAGTGTTCACCTGTTGATGATAAAATACCTGCACCATA 244
QY 268 TGCCTTTCTTCCTTCATGCTTTTCAATAAGTCCGCTTCAACAGTAACCAAAAGCA 324
DB 245 TGCCTAAATTGATACCTCTTTACATAATCCAAATTCAACTGTAACCAATAACA 301

RESULT 32
BJ399159

```

```

/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone_lib="Dictyostelium discoideum cDNA library, SF"
/sex="mat A"
/dev_stage="Slug stage"
BASE COUNT   227 a 101 c 119 g 227 t 1 others
ORIGIN
Query Match 4.1%; Score 45; DB 13; Length 675;
Best Local Similarity 61.5%; Pred. No. 0.81;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 208 TACGTTATCAATGAAGCGGTGCCAAGTTCCTGAGGAGAACTGATAAGACGGCTCCATA 267
DB 188 TGCCTTATCAGTTAAAGAGTGTTCACCTGTTGATGATAAAATACCTGCACCATA 247
QY 268 TGCCTTTCTTCCTTCATGCTTTTCAATAAGTCCGCTTCAACAGTAACCAAAAGCA 324
DB 248 TGCCTAAATTGATACCTCTTTACATAATCCAAATTCAACTGTAACCAATAACA 304

RESULT 31
BJ343093
LOCUS          680 bp      mRNA      linear      EST 07-MAR-2002
DEFINITION    Dictyostelium discoideum cDNA library, AF Dictyostelium
VERSION       BJ343093
ACCESSION    BJ343093
KEYWORDS     Dictyostelium discoideum.
SOURCE       Dictyostelium discoideum.
ORGANISM     Dictyostelium discoideum.
REFERENCE    1 (bases 1 to 680)
AUTHORS      Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE        Full length cDNA of Dictyostelium discoideum at the aggregation
JOURNAL      stage
COMMENT      Unpublished (2002)
              Contact: Tadasu Shin-i
              Center For Genetic Resource Information
              National Institute of Genetics
              1111 Yata, Mishima, Shizuoka 411-8540, Japan
              Tel: 81-559-81-6856
              Fax: 81-559-81-6855
              Email: tshini@genes.nig.ac.jp.
FEATURES     source
              1..680
              /organism="Dictyostelium discoideum"
              /strain="AX4"
              /db_xref="taxon:44689"
              /clone_lib="Dictyostelium discoideum cDNA library, AF"
              /sex="mat A"
              /dev_stage="Aggregation stage"
BASE COUNT   225 a 99 c 122 g 232 t 2 others
ORIGIN
Query Match 4.1%; Score 45; DB 13; Length 680;
Best Local Similarity 61.5%; Pred. No. 0.81;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 208 TACGTTATCAATGAAGCGGTGCCAAGTTCCTGAGGAGAACTGATAAGACGGCTCCATA 267
DB 185 TGCCTTATCAGTTAAAGAGTGTTCACCTGTTGATGATAAAATACCTGCACCATA 244
QY 268 TGCCTTTCTTCCTTCATGCTTTTCAATAAGTCCGCTTCAACAGTAACCAAAAGCA 324
DB 245 TGCCTAAATTGATACCTCTTTACATAATCCAAATTCAACTGTAACCAATAACA 301

RESULT 32
BJ399159

```



```

VERSION      BJ400818.1  GI:19313735
KEYWORDS     EST.
SOURCE       Dictyostellium discoideum.
ORGANISM     Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.
REFERENCE    (bases 1 to 718)
AUTHORS     Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE       Full length cDNA of Dictyostellium discoideum at the slug stage
JOURNAL     Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshinigenes.nig.ac.jp.

FEATURES     source
             1..718
             /organism="Dictyostellium discoideum"
             /strain="AX4"
             /db_xref="taxon:44689"
             /clone_lib="dictyostellium discoideum cDNA library, SF"
             /sex="mat A"
             /dev_stage="Slug stage"
BASE COUNT   241 a 104 c 122 g 248 t 3 others
ORIGIN
Query Match 4.1%; Score 45; DB 13; Length 718;
Best Local Similarity 61.5%; Pred. No. 0.81;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 208 TAGCTTATCAATGAAGCGTGTCCAAAGTTCCTGAGGAGAACTGATAAGACGGCTCCATA 267
DB 198 TGTCTTATCAAGTAAAGAGTGTCCATTTACCTGTGATGATAAAATACCTGCACCATTA 257
QY 268 TGCCTTTCTTCCTTCATGGTTTTCATTAAGTCGCTTCAACAGTAACCAACCAAGCA 324
DB 258 TGCTCTAATTGTATCACCTTCTTTACATAATCCAATTCAACTGTAACCAATAACA 314

RESULT 36
LOCUS       BJ401195
DEFINITION BJ401195 Dictyostellium discoideum cDNA library, EST 10-MAR-2002
ACCESSION  BJ401195
VERSION     BJ401195.1  GI:19314112
KEYWORDS   Dictyostellium discoideum.
SOURCE     Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.
REFERENCE  1 (bases 1 to 722)
AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE     Full length cDNA of Dictyostellium discoideum at the slug stage
JOURNAL   Unpublished (2002)
COMMENT   Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshinigenes.nig.ac.jp.
            Location/Qualifiers
              1..722
              /organism="Dictyostellium discoideum"
              /strain="AX4"
              /db_xref="taxon:44689"
              /clone_lib="dictyostellium discoideum cDNA library, SF"
              /sex="mat A"
              /dev_stage="Slug stage"
BASE COUNT 243 a 105 c 122 g 249 t 3 others
ORIGIN

```

```

ORIGIN
Query Match 4.1%; Score 45; DB 13; Length 722;
Best Local Similarity 61.5%; Pred. No. 0.81;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 208 TAGCTTATCAATGAAGCGTGTCCAAAGTTCCTGAGGAGAACTGATAAGACGGCTCCATA 267
DB 237 TGTCTTATCAAGTAAAGAGTGTCCATTTACCTGTGATGATAAAATACCTGCACCATTA 296
QY 268 TGCCTTTCTTCCTTCATGGTTTTCATTAAGTCGCTTCAACAGTAACCAACCAAGCA 324
DB 297 TGCTCTAATTGTATCACCTTCTTTACATAATCCAATTCAACTGTAACCAATAACA 353

RESULT 37
LOCUS       BJ341261
DEFINITION BJ341261 Dictyostellium discoideum cDNA library, AF Dictyostellium
ACCESSION  BJ341261
VERSION     BJ341261.1  GI:19249623
KEYWORDS   Dictyostellium discoideum.
SOURCE     Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.
REFERENCE  1 (bases 1 to 724)
AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE     Full length cDNA of Dictyostellium discoideum at the aggregation
JOURNAL   Unpublished (2002)
COMMENT   Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshinigenes.nig.ac.jp.
            Location/Qualifiers
              1..724
              /organism="Dictyostellium discoideum"
              /strain="AX4"
              /db_xref="taxon:44689"
              /clone_lib="dictyostellium discoideum cDNA library, AF"
              /sex="mat A"
              /dev_stage="Aggregation stage"
BASE COUNT 243 a 106 c 124 g 251 t
ORIGIN
Query Match 4.1%; Score 45; DB 13; Length 724;
Best Local Similarity 61.5%; Pred. No. 0.81;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 208 TAGCTTATCAATGAAGCGTGTCCAAAGTTCCTGAGGAGAACTGATAAGACGGCTCCATA 267
DB 237 TGTCTTATCAAGTAAAGAGTGTCCATTTACCTGTGATGATAAAATACCTGCACCATTA 296
QY 268 TGCCTTTCTTCCTTCATGGTTTTCATTAAGTCGCTTCAACAGTAACCAACCAAGCA 324
DB 297 TGCTCTAATTGTATCACCTTCTTTACATAATCCAATTCAACTGTAACCAATAACA 353

RESULT 38
LOCUS       BJ343890
DEFINITION BJ343890 Dictyostellium discoideum cDNA library, AF Dictyostellium
ACCESSION  BJ343890
VERSION     BJ343890.1  GI:19214397
KEYWORDS   Dictyostellium discoideum.
SOURCE     Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.
REFERENCE  1 (bases 1 to 734)
AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE     Full length cDNA of Dictyostellium discoideum at the slug stage
JOURNAL   Unpublished (2002)
COMMENT   Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshinigenes.nig.ac.jp.
            Location/Qualifiers
              1..734
              /organism="Dictyostellium discoideum"
              /strain="AX4"
              /db_xref="taxon:44689"
              /clone_lib="dictyostellium discoideum cDNA library, AF"
              /sex="mat A"
              /dev_stage="Slug stage"
BASE COUNT 243 a 105 c 122 g 249 t 3 others
ORIGIN

```

```

Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.
1 (bases 1 to 734)
Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
Full length cDNA of Dictyostellium discoideum at the aggregation
stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
source
1..734
Location/Qualifiers
/organism="Dictyostellium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone_lib="dictyostellium discoideum cDNA library, AF"
/dev_stage="Aggregation stage"
BASE COUNT 247 a 105 c 126 g 256 t
ORIGIN
Query Match 4.1%; Score 45; DB 13; Length 734;
Best Local Similarity 61.5%; Pred. No. 0.81;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 208 TAGCTTATCAATGAAGCGTGTCCAAAGTTCCTGAGGAGAACTGATAAGACGGCTCCATA 267
1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||
Db 236 TGCCTTATCAGTTAAAAGTGTTCATTTCCACCTGTGTGATGATAAAATACCTGCACCATA 295
1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||
Qy 268 TGCCTTTCTCTTCATCGTTTCAATAAGTCCGCTTTCACAGTAACCAACCAAGCA 324
1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||
Db 296 TGCTCTAATTGTATCACCTTCTTTACATAATCAAAATTCACCTGTAACCAATAACA 352
1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||

RESULT 39
BJ399243 742 bp mRNA linear EST 10-MAR-2002
LOCUS Dictyostellium discoideum cDNA library, SF Dictyostellium
DEFINITION dictyostellium cDNA clone dds5c06 3', mRNA sequence.
ACCESSION BJ399243
VERSION BJ399243.1 GI:19312160
KEYWORDS Dictyostellium discoideum.
SOURCE Dictyostellium discoideum.
ORGANISM Dictyostellium discoideum.
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.
1 (bases 1 to 742)
Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
Full length cDNA of Dictyostellium discoideum at the slug stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
source
1..742
Location/Qualifiers
/organism="Dictyostellium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone_lib="dictyostellium discoideum cDNA library, SF"
/dev_stage="Slug stage"
BASE COUNT 249 a 105 c 126 g 261 t 1 others
ORIGIN
Query Match 4.1%; Score 45; DB 13; Length 742;

```

```

Best Local Similarity 61.5%; Pred. No. 0.81;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 208 TAGCTTATCAATGAAGCGTGTCCAAAGTTCCTGAGGAGAACTGATAAGACGGCTCCATA 267
1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||
Db 236 TGCCTTATCAGTTAAAAGTGTTCATTTCCACCTGTGTGATGATAAAATACCTGCACCATA 295
1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||
Qy 268 TGCCTTTCTCTTCATCGTTTCAATAAGTCCGCTTTCACAGTAACCAACCAAGCA 324
1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||
Db 296 TGCTCTAATTGTATCACCTTCTTTACATAATCAAAATTCACCTGTAACCAATAACA 352
1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||

RESULT 40
BJ347280 744 bp mRNA linear EST 06-MAR-2002
LOCUS Dictyostellium discoideum cDNA library, AF Dictyostellium
DEFINITION dictyostellium cDNA clone dda26k16 3', mRNA sequence.
ACCESSION BJ347280
VERSION BJ347280.1 GI:19217787
KEYWORDS Dictyostellium discoideum.
SOURCE Dictyostellium discoideum.
ORGANISM Dictyostellium discoideum.
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.
1 (bases 1 to 744)
Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
Full length cDNA of Dictyostellium discoideum at the aggregation
stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1..744
/organism="Dictyostellium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone_lib="dictyostellium discoideum cDNA library, AF"
/dev_stage="Aggregation stage"
BASE COUNT 249 a 105 c 127 g 262 t 1 others
ORIGIN
Query Match 4.1%; Score 45; DB 13; Length 744;
Best Local Similarity 61.5%; Pred. No. 0.81;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 208 TAGCTTATCAATGAAGCGTGTCCAAAGTTCCTGAGGAGAACTGATAAGACGGCTCCATA 267
1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||
Db 236 TGCCTTATCAGTTAAAAGTGTTCATTTCCACCTGTGTGATGATAAAATACCTGCACCATA 295
1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||
Qy 268 TGCCTTTCTCTTCATCGTTTCAATAAGTCCGCTTTCACAGTAACCAACCAAGCA 324
1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||
Db 296 TGCTCTAATTGTATCACCTTCTTTACATAATCAAAATTCACCTGTAACCAATAACA 352
1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||

RESULT 41
BJ401342 760 bp mRNA linear EST 10-MAR-2002
LOCUS Dictyostellium discoideum cDNA library, SF Dictyostellium
DEFINITION dictyostellium cDNA clone dds22019 3', mRNA sequence.
ACCESSION BJ401342
VERSION BJ401342.1 GI:19314259
KEYWORDS Dictyostellium discoideum.
SOURCE Dictyostellium discoideum.
ORGANISM Dictyostellium discoideum.
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.
1 (bases 1 to 760)
Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.

```

**TITLE** Full length cDNA of Dictyostellium discoideum at the slug stage  
**JOURNAL** Unpublished (2002)  
**COMMENT** Contact: Tadasu Shin-i  
 Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshini@genes.nig.ac.jp.

**FEATURES** Location/Qualifiers

**source**  
 1..760  
 /organism="Dictyostellium discoideum"  
 /strain="AX4"  
 /db\_xref="taxon:44689"  
 /clone\_lib="ddc22019"  
 /clone\_lib="Dictyostellium discoideum cDNA library, SF"  
 /sex="mat A"  
 /dev\_stage="slug stage"  
 257 a 109 c 127 g 266 t 1 others

**BASE COUNT** 257 a 109 c 127 g 266 t 1 others  
**ORIGIN**

Query Match 4.1%; Score 45; DB 13; Length 760;  
 Best Local Similarity 61.5%; Pred. No. 0.81;  
 Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

**Qy** 208 TACCTTATCAATGAAGCGTGTCCAGTCTCTGAGGAGAACTGATAAGAACGGCTCCATA 267  
**Db** 237 TGCCTTATCAATGAAGCGTGTCCAGTCTCTGAGGAGAACTGATAAGAACGGCTCCATA 296  
**Qy** 268 TGCCTTATCAATGAAGCGTGTCCAGTCTCTGAGGAGAACTGATAAGAACGGCTCCATA 324  
**Db** 297 TGCCTTATCAATGAAGCGTGTCCAGTCTCTGAGGAGAACTGATAAGAACGGCTCCATA 353

**RESULT 42**

**LOCUS** B371948 765 bp mRNA linear EST 08-MAR-2002  
**DEFINITION** B371948 Dictyostellium discoideum cDNA library, CF Dictyostellium  
 discoideum cDNA clone ddc10p12 3', mRNA sequence.

**ACCESSION** B371948  
**VERSION** B371948.1 GI:19281331  
**KEYWORDS** EST.  
**SOURCE** Dictyostellium discoideum.  
**ORGANISM** Dictyostellium discoideum.

**REFERENCE** 1 (bases 1 to 765)  
 Eukaryota; Mycetozoa; Dictyostelidia; Dictyostellium.  
 Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.

Full length cDNA of Dictyostellium discoideum at the culmination  
 stage

**TITLE** Unpublished (2002)  
**JOURNAL** Contact: Tadasu Shin-i  
**COMMENT** Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshini@genes.nig.ac.jp.

**FEATURES** Location/Qualifiers

**source**  
 1..765  
 /organism="Dictyostellium discoideum"  
 /strain="AX4"  
 /db\_xref="taxon:44689"  
 /clone\_lib="ddc10p12"  
 /clone\_lib="Dictyostellium discoideum cDNA library, CF"  
 /sex="mat A"  
 /dev\_stage="Culmination stage"  
 258 a 113 c 126 g 268 t

**BASE COUNT** 258 a 113 c 126 g 268 t  
**ORIGIN**

Query Match 4.1%; Score 45; DB 13; Length 765;  
 Best Local Similarity 61.5%; Pred. No. 0.81;  
 Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

**Qy** 208 TACCTTATCAATGAAGCGTGTCCAGTCTCTGAGGAGAACTGATAAGAACGGCTCCATA 267  
**Db** 237 TGCCTTATCAATGAAGCGTGTCCAGTCTCTGAGGAGAACTGATAAGAACGGCTCCATA 296  
**Qy** 268 TGCCTTATCAATGAAGCGTGTCCAGTCTCTGAGGAGAACTGATAAGAACGGCTCCATA 324  
**Db** 297 TGCCTTATCAATGAAGCGTGTCCAGTCTCTGAGGAGAACTGATAAGAACGGCTCCATA 353

**RESULT 43**

**LOCUS** CNS0072R 988 bp DNA linear GSS 03-JUN-1999  
**DEFINITION** Drosophila melanogaster genome survey sequence TET3 end of BAC #  
 BACR14E13 of RPCI-98 library from Drosophila melanogaster (fruit  
 fly) genomic survey sequence.  
**ACCESSION** AL066743  
**VERSION** AL066743.1 GI:4945206  
**KEYWORDS** GSS.  
**SOURCE** Drosophila melanogaster.  
**ORGANISM** Drosophila melanogaster.  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.

**REFERENCE** 1 (bases 1 to 988)  
**AUTHORS** Genoscope.  
**JOURNAL** Direct Submission  
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)

**COMMENT** Determination of this BAC-end sequence was carried out as part of a  
 collaboration with the Berkeley Drosophila Genome Project (BDGP).  
 The BDGP is constructing a physical map of the Drosophila  
 melanogaster genome using these BACs. For further information  
 please see <http://www.fruitfly.org> The BDGP Drosophila  
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
 Aaron Mammos in the Roswell Park Cancer Institute in Buffalo,  
 NY. The library is named RPCI-98 and was constructed by partial  
 EcoRI digestion of Drosophila DNA provided by the BDGP from the  
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
 P1 and EST libraries. A more detailed description of the library  
 and how to order individual BAC clones, the entire library, or  
 filters for hybridization from the BACPAC Resource Center can be  
 found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

**FEATURES** Location/Qualifiers  
 1..988  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone="BACR14E13"  
 /clone\_lib="RPCI-98"  
 /note="end : TET3"

**BASE COUNT** 462 a 175 c 86 g 159 t 106 others  
**ORIGIN**

Query Match 4.1%; Score 44.6; DB 17; Length 988;  
 Best Local Similarity 39.3%; Pred. No. 1;  
 Matches 107; Conservative 31; Mismatches 134; Indels 0; Gaps 0;

**Qy** 623 AAATGTGTTTCTAATTCGAGAAATTTAATGACTGCTGTGATGCTTAAGAAATCAGAAAGG 682  
**Db** 540 AACTYATCCAAAAMWMAAAAMWMTTAAATAAAAAAACCTTAAAAAATAAAAAA 599  
**Qy** 683 AGACCGAAGCTTCTAAGTAATCAAGAAAAAATCTTGACAAATAGCTTTTCCATAGAAA 742  
**Db** 600 AAACCTAAAHTCTCAAAAAAATAAAAAATCTYTAATAAAAAAMCTTTCMACAYACA 659  
**Qy** 743 AACCGTGAGGAAGGAGTCGATACCAATAATTCGGATTAAAAAAGGAGCTTACACA 802  
**Db** 660 WTMAAAWHCCAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 719  
**Qy** 803 GCCATATCAGTGAGCAACCTCCCTTTGATTCCTTTGATATGACAGCAAGATCAACA 862  
**Db** 720 YTTATWTCYTAATTYTATATCAAAATWTHNTCAHAAATATTAATAAAAAAATAAAAA 779





## RESULT 46

AZ550349  
LOCUS ENTEM79TF Entamoeba histolytica sheared DNA linear GSS 14-NOV-2000  
DEFINITION genomic, DNA sequence.  
ACCESSION AZ550349  
VERSION AZ550349.1 GI:11175650  
KEYWORDS GSS.  
SOURCE Entamoeba histolytica.  
ORGANISM Entamoeba histolytica.  
REFERENCE 1 (bases 1 to 896)  
AUTHORS Loftus B., Van Aken S. and Fraser C.  
TITLE Determination of clone end sequences from Entamoeba histolytica  
JOURNAL HMI:IMSS sheared DNA library  
COMMENT Unpublished (2000)  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjloftus@tigr.org  
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
DNA library

Seq primer: M13-Forward  
Class: shotgun  
High quality sequence start: 36  
High quality sequence stop: 844.  
FEATURES  
source  
1..896  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica sheared DNA"  
/note="Vector: PHOS1; Site\_1: Bst I; Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described by Clark and Diamond (Clark,  
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
method for isolate identification. Exp. Parasitol.  
77:450.). The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The v + i method used for  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for  
whole genome shotgun sequencing projects. In Genome  
Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Barell, Oxford University Press, 1999)."

BASE COUNT  
ORIGIN

Query Match 4.0%; Score 43.8; DB 17; Length 896;  
Best Local Similarity 52.5%; Pred. No. 1.7;  
Matches 96; Conservative 0; Mismatches 87; Indels 0; Gaps 0;  
Qy 532 TCGCATTTACAGAGCAAGTGGGAAAGTACCGTCTCGACACAGACAGCAAGTATTGATGGG 591  
Db 430 TCAAAATTACAAAAGAAATTTATTTCAATTATTTGTTGAAAAACAAAAGAAATTTAAAGGAA 489  
Qy 592 AGCTACAAATCTGAAACGGGATAATAGGAAAAATGTGTTCTTAATTCGAAGAATTTAAT 651  
Db 490 TTTTAAAGAAGATGGAACAAGAATAAGAAACATGAATTAATTAATTAATAAAGAAAGGAA 549  
Qy 652 GACTGCTTGATGCTCTAAGAAATCAGAAAGGAGACCGCAAGCTTCTTAAGTAATCAAGAAA 711  
Db 550 GAAAAATGGATAGATTAAAAAACAAGAAACAATGGAACACACAAAAGATGGAGAA 609  
Qy 712 AAA 714  
Db 610 AAA 612

## RESULT 47

AZ527921/c  
LOCUS ENTCG177R Entamoeba histolytica sheared DNA linear GSS 03-NOV-2000  
DEFINITION genomic, DNA sequence.  
ACCESSION AZ527921  
VERSION AZ527921.1 GI:11080094  
KEYWORDS GSS.  
SOURCE Entamoeba histolytica.  
ORGANISM Entamoeba histolytica.  
REFERENCE 1 (bases 1 to 899)  
AUTHORS Loftus B., Van Aken S. and Fraser C.  
TITLE Determination of clone end sequences from Entamoeba histolytica  
JOURNAL HMI:IMSS sheared DNA library  
COMMENT Unpublished (2000)  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjloftus@tigr.org  
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
DNA library

Seq primer: M13-Reverse  
Class: shotgun  
High quality sequence start: 15  
High quality sequence stop: 808.  
FEATURES  
source  
1..899  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica sheared DNA"  
/note="Vector: PHOS1; Site\_1: Bst I; Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described by Clark and Diamond (Clark,  
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
method for isolate identification. Exp. Parasitol.  
77:450.). The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The v + i method used for  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for  
whole genome shotgun sequencing projects. In Genome  
Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Barell, Oxford University Press, 1999)."

BASE COUNT 208 a 151 c 64 g 476 t  
ORIGIN

Query Match 4.0%; Score 43.8; DB 17; Length 899;  
Best Local Similarity 52.5%; Pred. No. 1.7;  
Matches 96; Conservative 0; Mismatches 87; Indels 0; Gaps 0;  
Qy 532 TCGCATTTACAGAGCAAGTGGGAAAGTACCGTCTCGACACAGACAGCAAGTATTGATGGG 591  
Db 554 TCAAAATTACAAAAGAAATTTATTTCAATTATTTGTTGAAAAACAAAAGAAATTTAAAGGAA 495  
Qy 592 AGCTACAAATCTGAAACGGGATAATAGGAAAAATGTGTTCTTAATTCGAAGAATTTAAT 651  
Db 494 TTTTAAAGAAGATGGAACAAGAATAAGAAACATGAATTAATTAATTAATAAAGAAAGGAA 435  
Qy 652 GACTGCTTGATGCTCTAAGAAATCAGAAAGGAGACCGCAAGCTTCTTAAGTAATCAAGAAA 711  
Db 434 GAAAAATGGATAGATTAAAAAACAAGAAACAATGGAACACACAAAAGATGGAGAA 375  
Qy 712 AAA 714  
Db 374 AAA 372

RESULT 48  
 AZ687856/c  
 LOCUS ENTL72TF Entamoeba histolytica Sheared DNA Entamoeba histolytica  
 DEFINITION genomic, DNA sequence.  
 ACCESSION AZ687856  
 VERSION AZ687856.1 GI:11825002  
 KEYWORDS GSS.  
 SOURCE Entamoeba histolytica.  
 ORGANISM Eukaryota; Entamoebidae; Entamoeba.  
 REFERENCE 1 (bases 1 to 914)  
 AUTHORS Loftus, B., Van Aken, S. and Fraser, C.  
 TITLE Determination of clone end sequences from Entamoeba histolytica  
 JOURNAL HMI:IMSS sheared DNA library  
 COMMENT Unpublished (2000)  
 Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: bjloftus@tigr.org  
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
 DNA library  
 Seq primer: M13-Forward  
 Class: shotgun  
 High quality sequence start: 7  
 High quality sequence stop: 869.  
 Location/Qualifiers  
 1. 914  
 /organism="Entamoeba histolytica"  
 /strain="HMI:IMSS"  
 /db\_xref="taxon:5759"  
 /clone\_lib="Entamoeba histolytica Sheared DNA"  
 /note="Vector: PHOS1; Site 1: Bst I; Constructed at The  
 Institute for Genomic Research (TIGR), Rockville, MD.  
 Genomic DNA isolated from broth cultures of E. histolytica  
 using a method described by Clark and Diamond (Clark,  
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
 method for isolate identification. Exp. Parasitol.  
 77:450.). The DNA was mechanically sheared to give a  
 tight size distribution (~2 kb). The v + i method used for  
 the library construction is described in detail in Smith,  
 H.O. and Venter, J.C. (Making small insert libraries for  
 whole genome shotgun sequencing projects. In Genome  
 Sequencing: A Practical Approach, eds. M. Vaudin and B.  
 Borell, Oxford University Press, 1999)."

## FEATURES

source

BASE COUNT 209 a 155 c 85 g 465 t  
 ORIGIN  
 Query Match 4.0%; Score 43.8; DB 17; Length 914;  
 Best Local Similarity 52.5%; Pred. No. 1.7;  
 Matches 96; Conservative 0; Mismatches 87; Indels 0; Gaps 0;  
 QY 532 TCGCATTACAGAGGCAATGGAAAGTAAACGGTCTCCACACAGACAAAGTATTCATGGGG 591  
 Db 260 TCAAAATACAAAAGAAATTTATTTCAATTAATTTGTAACAAAGAAATTTAAAGGAA 201  
 QY 592 AGCTACAATCTCGAAGGGATATAGGAAAATCTGTTTCTAATTCGAAGAATTTAAT 651  
 Db 200 TTTTGAAGAAGTGGCAAGAAGAAATAGAAACATGAATTAATTAATTAATAAAGGAA 141  
 QY 652 GACTGCTTGATGCTTAAGAATCAGAAAGGAGACCGGAAGCTTCTAAGTAATCAAGAAA 711  
 Db 140 GAAAAATGGATAGTATTAATAAACAACAAAGAAACAAATGGAACACAAAAAGTGGAGAA 81  
 QY 712 AAA 714  
 Db 80 AAA 78

RESULT 49

AZ690917  
 LOCUS ENTHV65TR Entamoeba histolytica Sheared DNA Entamoeba histolytica  
 DEFINITION genomic, DNA sequence.  
 ACCESSION AZ690917  
 VERSION AZ690917.1 GI:11828063  
 KEYWORDS GSS.  
 SOURCE Entamoeba histolytica.  
 ORGANISM Eukaryota; Entamoebidae; Entamoeba.  
 REFERENCE 1 (bases 1 to 924)  
 AUTHORS Loftus, B., Van Aken, S. and Fraser, C.  
 TITLE Determination of clone end sequences from Entamoeba histolytica  
 JOURNAL HMI:IMSS sheared DNA library  
 COMMENT Unpublished (2000)  
 Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: bjloftus@tigr.org  
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
 DNA library  
 Seq primer: M13-Reverse  
 Class: shotgun  
 High quality sequence start: 85  
 High quality sequence stop: 895.  
 Location/Qualifiers  
 1. 924  
 /organism="Entamoeba histolytica"  
 /strain="HMI:IMSS"  
 /db\_xref="taxon:5759"  
 /clone\_lib="Entamoeba histolytica Sheared DNA"  
 /note="Vector: PHOS1; Site 1: Bst I; Constructed at The  
 Institute for Genomic Research (TIGR), Rockville, MD.  
 Genomic DNA isolated from broth cultures of E. histolytica  
 using a method described by Clark and Diamond (Clark,  
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
 method for isolate identification. Exp. Parasitol.  
 77:450.). The DNA was mechanically sheared to give a  
 tight size distribution (~2 kb). The v + i method used for  
 the library construction is described in detail in Smith,  
 H.O. and Venter, J.C. (Making small insert libraries for  
 whole genome shotgun sequencing projects. In Genome  
 Sequencing: A Practical Approach, eds. M. Vaudin and B.  
 Borell, Oxford University Press, 1999)."

## FEATURES

source

BASE COUNT 482 a 63 c 169 g 210 t  
 ORIGIN  
 Query Match 4.0%; Score 43.8; DB 17; Length 924;  
 Best Local Similarity 52.5%; Pred. No. 1.7;  
 Matches 96; Conservative 0; Mismatches 87; Indels 0; Gaps 0;  
 QY 532 TCGCATTACAGAGGCAATGGAAAGTAAACGGTCTCCACACAGACAAAGTATTCATGGGG 591  
 Db 116 TCAAAATACAAAAGAAATTTATTTCAATTAATTTGTAACAAAGAAATTTAAAGGAA 175  
 QY 592 AGCTACAATCTCGAAGGGATATAGGAAAATCTGTTTCTAATTCGAAGAATTTAAT 651  
 Db 176 TTTTGAAGAAGTGGCAAGAAGAAATACAAACATGAATTAATTAATTAATAAAGGAA 235  
 QY 652 GACTGCTTGATGCTTAAGAATCAGAAAGGAGACCGGAAGCTTCTAAGTAATCAAGAAA 711  
 Db 236 GAAATGGATAGTATTAATAAACAACAAAGAAACAAATGGAACACAAAAAGTGGAGAA 295  
 QY 712 AAA 714  
 Db 296 AAA 298  
 RESULT 50  
 AZ549336

LOCUS	A2549336	944 bp	DNA	linear	GSS 14-NOV-2000			
DEFINITION	ENTED22TRB Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.							
ACCESSION	A2549336							
VERSION	A2549336.1	GI:11173846						
KEYWORDS	GSS.							
SOURCE	Entamoeba histolytica.							
ORGANISM	Entamoeba histolytica							
REFERENCE	Eukaryota; Entamoebidae; Entamoeba.							
AUTHORS	1 (bases 1 to 944)							
TITLE	Loftus, B., Van Aken, S. and Fraser, C.							
	Determination of clone end sequences from Entamoeba histolytica							
JOURNAL	HM1:IMSS sheared DNA library							
COMMENT	Unpublished (2000)							
	Contact: Brendan J Loftus							
	Department of Eukaryotic Genomics							
	The Institute for Genomic Research							
	9712 Medical Center Dr., Rockville, MD 20850, USA							
	Tel: 301 838 0208							
	Fax: 301 838 3543							
	Email: bjlloftus@tigr.org							
	Clones are derived from the Entamoeba histolytica HM1:IMSS sheared							
	DNA library							
	Seq primer, M13-Reverse							
	Class: shotgun							
	High quality sequence start: 15							
	High quality sequence stop: 907.							
FEATURES	Location/Qualifiers							
source	1..944							
	/organism="Entamoeba histolytica"							
	/strain="HM1:IMSS"							
	/db_xref="taxon:5759"							
	/clone_lib="Entamoeba histolytica Sheared DNA"							
	/note="Vector: PHO31; Site_1: Bst I; Constructed at The							
	Institute for Genomic Research (TIGR), Rockville, MD.							
	Genomic DNA isolated from broth cultures of E. histolytica							
	using a method described by Clark and Diamond (Clark,							
	C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a							
	method for isolate identification. Exp. Parasitol.							
	77:450.). The DNA was mechanically sheared to give a							
	tight size distribution (~2 kb). The v + i method used for							
	the library construction is described in detail in Smith,							
	H.O. and Venter, J.C. (Making small insert libraries for							
	whole genome shotgun sequencing projects. In Genome							
	Sequencing: A Practical Approach, eds. M. Vaudin and B.							
	Barell, Oxford University Press, 1999)."							
BASE COUNT	500 a	60 c	162 g	222 t				
ORIGIN								

	Query Match	4.0%;	Score 43.8;	DB 17;	Length 944;
	Best Local Similarity	52.5%;	Prod. No. 1.7;		
	Matches 96;	Conservative	0;	Mismatches 87;	Indels 0; Gaps 0;
QY	532	TGCGATTACAGAGCGCAATGGGAAAGTACCGTCTCGAACAGAGACAGATATTGATGGG	591		
Db	376	TCAATTACAAAGAAGATTATTTTCAATTATGTTGAAAACCAAGAAGATTTAAAAAGAA	435		
QY	592	AGCTACAACTCTGAACGGGATATAGAAAAANTGTTCCTAATTCGAAGAAATTTAAT	651		
Db	436	TTTTAAGAAGATGGCAACAGAAATAGAAACATGAATTAATTAATTTAAAAAAGAAGAA	495		
QY	652	GACTGCTTGATGCTTAAGAAATCAGAAAGGAGACCGAAAGCTTCTAAGTAATCAAGAA	711		
Db	496	GAATAATGGATAGATTAAAAAACAAGAAAGAAACAATGGAAAAACAAGAAATGGAGAA	555		
QY	712	AAA 714			
Db	556	AAA 558			

Search completed: January 9, 2003, 11:40:55  
Job time : 1961 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 05:32:32 ; Search time 251 Seconds  
(without alignments)  
9743.706 Million cell updates/sec

Title: us-09-438-185a-1\_COPY\_1199590\_1200675

Perfect score: 1086  
Sequence: 1 ttggcaagtagtacctcaaac.....gtctctcgtagtagtgac 1086

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 150 summaries

Database : N\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	1086	100.0	1089	24	ABL91284
C 2	46.6	4.3	1642	21	AAA91624
C 3	43.4	4.0	6089	24	ABN80254
C 4	41.4	3.8	817	21	AAA91625
C 5	41	3.8	868	20	AA217250
C 6	41	3.8	9504	24	ABK28408
7	40.6	3.7	5152	24	ABL92307
8	40.6	3.7	5152	24	ABL49374
9	40.6	3.7	37184	24	ABQ67077
-----					
C 1	1086	100.0	1089	24	ABL91284
C 2	46.6	4.3	1642	21	AAA91624
C 3	43.4	4.0	6089	24	ABN80254
C 4	41.4	3.8	817	21	AAA91625
C 5	41	3.8	868	20	AA217250
C 6	41	3.8	9504	24	ABK28408
7	40.6	3.7	5152	24	ABL92307
8	40.6	3.7	5152	24	ABL49374
9	40.6	3.7	37184	24	ABQ67077

10	40.6	3.7	640681	24	ABA92787
11	40	3.7	6740	24	ABL33147
12	39.6	3.6	17419	22	AAS45393
13	39.6	3.6	17419	24	ABL33295
14	39.6	3.6	17419	24	ABK28238
C 15	39.6	3.6	99629	22	AAF28550
C 16	39.6	3.6	8895	24	ABL32333
C 17	38.6	3.6	326	23	ABV61126
18	38.4	3.5	18997	24	ABK33948
19	38.4	3.5	18997	24	ABL32570
C 20	38.4	3.5	50000	24	AAD26400
C 21	38.4	3.5	83391	24	ABQ67094
22	38.4	3.5	7667	22	AAS46353
23	38.2	3.5	9415	24	ABU32294
C 24	38.2	3.5	15373	24	ABU32467
C 25	38.2	3.5	47108	24	ABK31510
C 26	38.2	3.5	17213	24	ABL33483
27	38	3.5	113515	24	ABL34174
28	38	3.5	7690	24	ABL33122
C 29	37.8	3.4	870	21	AAC40730
C 30	37.4	3.4	6237	24	ABL32359
31	37.4	3.4	7644	24	ABL32530
32	37.4	3.4	522	21	AAC36518
C 33	37.2	3.4	869	21	AAC51121
C 34	37.2	3.4	3152	22	AAH54692
35	37.2	3.4	3345	22	AAH54004
36	37.2	3.4	6113	24	ABL32802
37	37.2	3.4	7190	24	ABL33720
38	37.2	3.4	619	23	ABV52129
C 39	37	3.4	1134	20	AAZ31005
40	37	3.4	1134	22	AAH34735
41	37	3.4	1134	22	AAH32611
C 42	36.8	3.4	1724	22	AAH32609
C 43	36.8	3.4	1724	22	AAH32609
44	36.8	3.4	1761	22	AAH32609
45	36.8	3.4	1761	22	AAH32609
46	36.8	3.4	7857	24	ABQ67075
C 47	36.8	3.4	294528	24	ABQ67075
C 48	36.6	3.4	2098	13	AAH07351
C 49	36.6	3.4	2098	13	AAH07351
C 50	36.6	3.4	2333	22	AAH17997
51	36.6	3.4	5276	24	ABL32151
52	36.6	3.4	7348	22	AAS46335
53	36.4	3.4	1241	22	AAD09807
C 54	36.4	3.4	6239	24	ABL70584
C 55	36.4	3.4	6239	24	ABL70584
C 56	36.4	3.4	6239	24	ABL70584
57	36.4	3.4	8962	24	ABL31387
58	36.4	3.4	15282	24	ABL32687
59	36.4	3.4	15282	24	ABL70189
60	36.4	3.4	21354	22	AAS61139
C 61	36.4	3.4	319630	24	AAS46815
C 62	36.2	3.3	373	23	ABQ67194
63	36.2	3.3	444	22	AAH69056
C 64	36.2	3.3	470	22	AAH71919
65	36.2	3.3	511	22	AAH70719
C 66	36.2	3.3	580	24	ABK44560
C 67	36.2	3.3	585	24	ABN66198
C 68	36.2	3.3	1985	20	AAZ32066
C 69	36.2	3.3	1985	22	AAC90323
C 70	36.2	3.3	1985	22	AAC89555
C 71	36.2	3.3	1985	24	ABK87716
C 72	36.2	3.3	6020	24	ABL49308
73	36.2	3.3	6020	24	ABL32199
74	36.2	3.3	6070	24	ABQ67129
75	36.2	3.3	6070	24	ABL70371
C 76	36.2	3.3	6070	24	ABL33678
C 77	36.2	3.3	6070	24	ABL34578
78	36.2	3.3	10988	24	ABL70200
C 79	36.2	3.3	10988	24	AAS61158
80	36.2	3.3	10988	24	ABK31245
81	36.2	3.3	17137	24	ABL32191
82	36.2	3.3	18998	22	AAS36452

Buchnera sp. genom  
Human immune syste  
Chemically pretrea  
Human immune syste  
DNA transcription  
Genomic fragment #  
Human immune syste  
Human prostate exp  
Human DNA for stag  
Human immune syste  
Human glutamate re  
Human GRM3 gene fr  
Human angio genesis  
Tumour suppressor  
Human immune syste  
Human immune syste  
Signal transductio  
Human immune syste  
Human immune syste  
Arabidopsis thalia  
Human immune syste  
Arabidopsis thalia  
S. epidermidis gen  
Human immune syste  
Human prostate exp  
Partial dnan gene.  
Nucleotide sequenc  
Human secreted pro  
Human immune/haema  
Human immune/haema  
Human angio genesis  
Listeria monocytog  
Human cDNA clone  
Homo sapiens solub  
Human cDNA sequenc  
Human immune syste  
Tumour suppressor  
Arabidopsis thalia  
Chemically treated  
Human gene regulat  
Signal transductio  
Human immune syste  
Chemically treated  
Human gene regulat  
Listeria innocua c  
Tumour suppressor  
Human prostate exp  
Human cervical can  
Human cervical can  
cDNA encoding colo  
Streptococcus poly  
Human METH2 relate  
U31814 cDNA clone.  
Human histone deac  
Human cDNA encodin  
Human polynucleoti  
Human immune syste  
Human angio genesis  
Chemically treated  
Human immune syste  
Human metastasis a  
Chemically treated  
Human gene regulat  
Signal transductio  
Human immune syste  
Human cardiovascular





Db 1464 TCTTCAAAATTTCTAGTATAGAAATACGCTGACGTGGAAGAGTGAATAGACATCTTTGTC 1405  
 Qy 173 CGAATAATCTGATCCAAATTTCTAAAGGGAGAACAGTACGTTATCAATGAAGCGGTGCTCA 232  
 Db 1404 TCAACACAGGATCCGGATCAAAAGGAATAATGTTGCACCTACCTCAACGCGATGTTGC 1345  
 Qy 233 AGTTCTGTGAGGAAGTATAGAAAGCGCTCCATATGCTTTTCTTCTTCTATGTTTTCATCA 292  
 Db 1344 AACTCGGCAGCACTGTCAGAAAGTCTGCTCGGTATATCTTTAAATCTTTCATGATTTGAT 1285  
 Qy 293 ATAAGTCGCTTTCACAGCTAAACAAAGCAGCGGTACATACGATCATGATGCTTTGCT 352  
 Db 1284 CCATTTTCTTTACTGGAG-ATTCGCGAGCGTCATCAGACGAGACCAAAATCAATGGA 1226  
 Qy 353 AGGTTGTATGCTGTTGTTTCTAGGAGAGCTGTGACTTTTCTTATGACTTTAGTG 412  
 Db 1225 AAGAAGTAGAGTGTGCAAGCTTCTTCAAA-----TCTTCTCTGATGCTCCAGA 1174  
 Qy 413 AAGAGTCTCCCATGTTTATGAAATTCAGAAATCAGAAATCAGGAGTGTAGAACCCAGCAG 472  
 Db 1173 GAGCTAATCCATCTCTTGAGAACTGAGCAAAATCTGATCAGCGATAGACCCATG 1114  
 Qy 473 TGCCTTAAAGGTCTAGTATGATCAGGAGTTAAGGAGAAATATCTTTATCGAGAGTT 532  
 Db 1113 TGACCATGAGCTGCGGAACGCTGCTGTTCTCGAGTGTAATAATGGATCGGCGATGAGG 1054  
 Qy 533 CGCATTCACAGAGCAATGGAAAGTAAAGTTCCTGTCGAACAGAGACAAAGTATTTGATGGGA 592  
 Db 1053 CGAACGCTATTGATGTCAGAAAGACAGCATATGCAAGACCTGCCAAGAAATCACGAGCT 994  
 Qy 593 GCTCAATTCCTGAAACGGGATATAGGAAATATGTTTCTTAAATTCGAGAAATTTAATG 652  
 Db 993 GATAAGTATCCGCGACTGGGCGAACACGGAATCCAGTTTCTGCTTTTCAAAACCTTGCG 934  
 Qy 653 ACTGCTT 659  
 Db 933 ATATCTT 927

RESULT 3  
 ABN80254/C  
 ID ABN80254 standard; DNA; 6089 BP.  
 AC ABN80254;  
 DT 15-JUL-2002 (first entry)  
 DE Human chemically modified disease associated gene SEQ ID NO 271.  
 KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;  
 KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;  
 KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;  
 KW antidiabetic; cytostatic; anticonvulsant; ds.  
 OS Homo sapiens.  
 OS Synthetic.  
 PN WO200200927-A2.  
 PD 03-JAN-2002.  
 PE 02-JUL-2001; 2001WO-EP07536.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 PR (EPIG-) EPIGENOMICS AG.  
 PA Olek A, Piepenbrock C, Berlin K;  
 PI WPT; 2002-130908/17.  
 DR Novel nucleic acid useful for diagnosis and therapy of diseases  
 XX

PT associated with development genes such as diabetes, comprises a  
 PT sequence of a segment of chemically pretreated DNA of genes associated  
 PT with development -  
 XX Claim 1: SEQ ID NO 271; 27pp; English.  
 XX The invention relates to a nucleic acid (I) comprising a sequence at  
 CC least 18 bases in length of a segment of chemically pretreated DNA (II)  
 CC of genes associated with development selected from 87 genes listed in  
 CC the specification such as ACCPN, ADFN, or APDI and comprising one of 350  
 CC sequences (ABN79984-ABN80333) or their complements. The invention is  
 CC useful for the diagnosis or therapy of diseases associated with  
 CC genes (HOX) like diabetes, cancer, apoptosis related diseases, syndromes  
 CC associated with congenital heart disease, epilepsy, diseases related to  
 CC histone deacetylation, Curarino syndrome, diseases related with the  
 CC development of the brain and limb girdle muscular dystrophy and dwarfism.  
 CC Oligomers specific to each of the genes are useful for detecting the  
 CC methylation state of all CpG dinucleotides within the 350 sequences or  
 CC (II) and their complementary sequences, as primer oligonucleotides for  
 CC the amplification of the 350 sequences, (II) and/or their complements and  
 CC as oligomer probes for detecting the cytosine methylation state and/or  
 CC single nucleotide polymorphisms (SNPs).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification but is based on sequence information supplied to Berwent by  
 CC the European Patent Office.  
 XX Sequence 6089 BP; 1262 A; 219 C; 1585 G; 3023 T; 0 other;  
 SQ

Query Match 4.0%; Score 43.4; DB 24; Length 6089;  
 Best Local Similarity 45.8%; Pred. No. 0.24;  
 Matches 149; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

Qy 621 AAAAATGTGTTCTTAATCGAAGAATTTAATGACTGCTTGATGCTTAAGAAATCAGAAA 680  
 Db 1350 AAAAAAATATTAACACCAAAATATATATACACAAATAAATACTATTAACACAAAAA 1291  
 Qy 681 GGAGACCGAAGCTTCTTAAGTATACAGAAAAATCTTGGACAATAGCTTTCCATAGAG 740  
 Db 1290 AACTCTCAACCAATCTCAAAAAATCAAAAAATCTCTTAAAAAATATTCCTAAAT 1231  
 Qy 741 AAAACCGTGAGGAAGAGTCGATACCAATAATCCGATTAAAAAAGGAGTCTTTACAA 800  
 Db 1230 AAAATCTTAAAAAATAATTAATTAACAAATAAATTTTAAACCAAAAAAACAA 1171  
 Qy 801 CAGCCATATCAGTGGAGCAACCTTCCCTTTGATTTCTTTGATTAATGACAGAGATGAACAA 860  
 Db 1170 AACCAAGAAATTCAAAATACCTAAATATTAACAGTTTAAAAAATAAATAATATACAAA 1111  
 Qy 861 AGAGTAAGTTTTTGAATAATCTTAAATAGAAATGCAATTTGTGTCGAGCTAAAGCTTGT 920  
 Db 1110 AAAAACTCAATAAAAAATCTAAAAACACAAAAACATATTAATAAAAAATAACAAATATTA 1051  
 Qy 921 TCTCTTTATTTCTTCTTTTGTAGAA 945  
 Db 1050 TAATCAAAATTTTATTTCAAAAAAT 1026

RESULT 4  
 AAA91625/C  
 ID AAA91625 standard; cDNA; 817 BP.  
 AC AAA91625;  
 DT 12-JAN-2001 (first entry)  
 DE Caenorhabditis elegans cod-5 knockout cDNA.  
 KW Caenorhabditis elegans; cod-5; daf-18; insulin signalling pathway;  
 KW age-1; insulin receptor; PI 3-kinase; PKB kinase;  
 KW PTEN lipid phosphatase; antidiabetic; anorectic; obesity; diabetes; ss.  
 OS Caenorhabditis elegans.





CC encoded protein); and to detect differences in expression levels between  
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to  
 CC identify a genetic predisposition or susceptibility to a disease such as  
 CC cancer). The polynucleotides of the invention are especially used in the  
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
 CC and lung cancer. The polynucleotides can also be used to screen for  
 CC peptide analogues and antagonists.

XX  
 SQ Sequence 868 BP; 148 A; 148 C; 32 G; 284 T; 256 other;  
 Query Match 3.8%; Score 41; DB 20; Length 868;  
 Best Local Similarity 33.1%; Pred. No. 0.51; Indels 0; Gaps 0;  
 Matches 89; Conservative 0; Mismatches 180;  
 QY 446 AATGAGGAGTGTAGAACCAAGCAGCGTCCCTAAAGGTCATGGATGAGATCAGGAGTT 505  
 DB 575 AGCNGGANGANNAATTAANNAAGNNGANGAANNAANNNTGAGANNAANAGAGGNT 516  
 QY 506 AAGGAGAAATTTCTTTATCGAGAGTTCGCATTTACAGAGCAATGGGAAGTAACGGTCC 565  
 DB 515 TAGGAGNNAANANANANGAGNNNNNNNTAGNNANNTAAATGGNAAANTGGANANN 456  
 QY 566 TGCAACAGAGACAAGTATTGTATGGGGAGCTACAAATCTCGAAACGGGATTAATAGGAAAA 625  
 DB 455 ANTWAGGNANAGNTNGAGNNNGAGGNATGANANATTGANAGAAANTNANNNNNNA 396  
 QY 626 TGTCTTTCTTAATTCGAGAAATTTAATGACTGCTGTGATGGTCTAAGAAATCAGAAAGGAGA 685  
 DB 395 NNNATNAATTAATGNNNGGANGANGANGAAGAAAAAANTTAAANNANTATNAATAAGN 336  
 QY 686 CCGAAAGCTTCTTAAGTAATCAAGAAAAA 714  
 DB 335 NNGAAGNTNGANGATGNTANANNA 307

RESULT 6  
 ABK28408  
 ID ABK28408 standard; DNA; 9504 BP.  
 XX  
 AC ABK28408;  
 DT 23-APR-2002 (first entry)  
 XX  
 DE DNA transcription associated complementary genomic DNA #141.  
 XX  
 KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;  
 KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;  
 KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;  
 KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;  
 KW immunological disorder; Werner's syndrome; developmental disorder;  
 KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;  
 KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;  
 KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;  
 KW angiodysplasia; congenital heart disease; HDR syndrome; gene therapy;  
 KW polyglutamine disorder; solid tumour.  
 XX  
 OS Unidentified.

XX  
 XX WO200192565-A2.  
 PN  
 PD 06-DEC-2001.  
 XX  
 PF 06-APR-2001; 2001WO-BP03973.  
 XX  
 PR 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX

DR  
 XX  
 PT New nucleic acids or oligomers, useful for diagnosing or treating  
 PT diseases associated with DNA transcription, e.g. immunological  
 PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid  
 PT tumours or cancer.  
 XX  
 PS Claim 1; SEQ ID NO 282; 32pp; English.  
 CC  
 CC The invention relates to a nucleic acid, which comprises a segment of the  
 CC chemically pretreated DNA of genes associated with DNA transcription from  
 CC one of 346 sequences, and an oligomer, in particular an oligonucleotide  
 CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical  
 CC to the chemically pretreated DNA of genes associated with DNA  
 CC transcription. The set of oligomer probes are useful for detecting the  
 CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)  
 CC in a chemically pretreated genomic DNA. The nucleic acids are useful for  
 CC diagnosing or treating diseases associated with DNA transcription  
 CC (particularly with the methylation status), e.g. adenosine deaminase  
 CC deficiency, viral infection, retroviral infection, Sezary syndrome,  
 CC haematological disorders, immunological disorders, Werner syndrome,  
 CC tuberculous, developmental disorders, psoriasis, Rieger's syndrome,  
 CC neurological disorders, neurodegenerative disorders, Waardenburg  
 CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial  
 CC infarction, hypertension, angiodysplasia, erythropoiesis, congenital heart  
 CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription  
 CC associated genomic DNA molecules of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format directly from the  
 CC European Patent Office.

SQ Sequence 9504 BP; 2775 A; 138 C; 2348 G; 4243 T; 0 other;  
 Query Match 3.8%; Score 41; DB 24; Length 9504;  
 Best Local Similarity 53.4%; Pred. No. 1.2;  
 Matches 86; Conservative 0; Mismatches 75; Indels 0; Gaps 0;  
 QY 862 GAGTAAGTCTTTTGAAGAAATCTAAATAGAAATGCGATTTGTCTGAGCTAAAGCTTCTT 921  
 DB 8911 GTGGATGATATATATTTGTTGTTATTTGTTTATTTGTTGTTGTTGTTGTTGTTT 8970  
 QY 922 CTCCTTTATTTTCTTTTGTAGAAATGCGGTAGTAGGAATATGGGTGCGAGTATGCACG 981  
 DB 8971 ATTGTTATTGTTGTTTATTTTATTTTATTTGAAAGAAAGTGGGTTAGGTTTAGAG 9030  
 QY 982 TTGGAGTATTGCTGCTTCTGGAAGAACAGGAGAAAGTGTATT 1022  
 DB 9031 ACATAGTGTGTTTGTGTTTGTAGGGGAATAGGATATTAATT 9071

RESULT 7  
 ABL92307  
 ID ABL92307 standard; DNA; 5152 BP.  
 XX  
 AC ABL92307;  
 DT 01-JUL-2002 (first entry)  
 XX  
 DE Chemically treated DNA repair gene fragment complementary to#58.  
 XX  
 KW DNA repair; cytosine methylation; PMS2L1; PMS2L2; PMS2L3;  
 KW PMS2; L4; PMS2L5; PMS2L6; MGMT; MSH2; NUDT1; TDG; INPPL1; RFC4;  
 KW DDT1L; FANCB; XRCC8; ataxia telangiectasia; aging; Bloom's syndrome;  
 KW Cockayne syndrome; Nijmegen breakage syndrome; Werner syndrome;  
 KW immunodeficiency; trichothiodystrophy; Fanconi's anaemia; solid tumour;  
 KW cancer; ds.  
 XX  
 OS Unidentified.  
 XX  
 PN WO200181622-A2.  
 XX  
 PD 01-NOV-2001.



Db 73 TGTGGGGTGGGAGGATATTTATTTAGTTAGTAAATTTGAAATTTGATGAGTCGTGA 132  
 Qy 644 AATTTAATGACTGCTGATGGTCTAAGAAATCAGAAAGGAGACCGAAGCTTCTAAGTAA 703  
 Db 133 GTATTTAGTTGGGTAATGAGTGTGAGATTTTGTATATAAATAAGATTTAGTTAGTATA 192  
 Qy 704 TCAAGAAAAATCTTGGCAATAGCTTTTCCATAGAGAAAAACCGTAGGAAAGGAGTCGA 763  
 Db 193 GATTTTAAAAATTTTAAATATTTTTTTTTTTTTTTAGGAAAGTGTGAAAAATAAGGAACG 252  
 Qy 764 TACCATAAATTCGATATAAAGGAGTGTACACACGCCATATCAGTCGGAGCAACCT 823  
 Db 253 ATTTTATTTTTTTTTTTAGAGATAAATTTTACATAAGATATTTAATGTAGGATTTG 312  
 Qy 824 TCCCTTTGATCTTTGATATCAGACAGAGATGACAAAGAGTAAGTTTGTGAAAAATCT 883  
 Db 313 AGAATGTTGATGTTTTAT-TTAAAGTAGGTGTTTTTATATATATATTTAGAAAAATTT 371  
 Qy 884 AAAATAGAAATGATTTGTCGAGCTAAAGCTTCTCTTTATTTTCCCTTTTCTAGTA 943  
 Db 372 TATTTATAAAATTTTATTTATTTTATTTTATTTTATTTATATATTTTATTTTCTTTA 431  
 Qy 944 ATGATTCGAGTAGTAAGAAATAGGGTCGAGTATGCACGCTGGAGTGATGGCTGTTCTGG 1003  
 Db 432 GTGTTTTTACGAAAAAATATATATTTTATTTTATTTTAAAGTCTTATAATTTTTTACT 491  
 Qy 1004 AAGAACAGGGAAGTGTGCTCAGCTTTA 1034  
 Db 492 TAATATTCGTAAGAAATTTTATAGATTTTA 522

RESULT 9  
 ABQ67077  
 ID ABQ67077 standard; DNA; 37184 BP.  
 XX AC ABQ67077;  
 XX DT 28-AUG-2002 (first entry)  
 XX DE Human angio genesis associated polynucleotide SEQ ID NO 107.  
 XX KW Human; angio genesis; methylation; eye disease; glaucoma; tumour;  
 KW inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcerds;  
 KW macular degeneration; inflammatory bowel disease; Crohn's disease;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antiarteriosclerotic; ds.  
 XX OS Homo sapiens.  
 XX PN WO200246454-A2.  
 XX PD 13-JUN-2002.  
 XX PF 06-DEC-2001; 2001WO-EP14320.  
 XX PR 06-DEC-2000; 2000DE-1061338.  
 XX PA (EPIG-) EPIGENOMICS AG.  
 XX PI Schacht O;  
 XX WPI; 2002-500450/53.  
 XX New nucleic acid fragments from chemically treated  
 PT angio genesis-associated genes, useful for determining methylation  
 PT status, e.g. in diagnosis or treatment of cancer  
 XX Claim 1; SEQ ID NO 107; 41pp + Sequence Listing; German.  
 XX The invention relates to a nucleic acid (I) comprising a segment of 18  
 CC bases of chemically pretreated DNA of angio genesis-associated genes (II)  
 CC having sequences (ABQ66971-ABQ67178) or their complements. (I), also

CC related oligomers, are used to evaluate the methylation status and/or  
 CC single-nucleotide polymorphisms, in angio genesis-related genes, for  
 CC diagnosis and treatment of eye diseases, proliferative retinopathy,  
 CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,  
 CC diabetic retinopathy, macular degeneration caused by neovascularisation,  
 CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and  
 CC Crohn's disease.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 37184 BP; 8680 A; 717 C; 9540 G; 18247 T; 0 other;

Query Match 3.7%; Score 40.6; DB 24; Length 37184;  
 Best Local Similarity 57.5%; Pred. No. 2.7;  
 Matches 73; Conservative 0; Mismatches 54; Indels 0; Gaps 0;  
 Qy 834 TCTTTGATAATGACAGAGATGACAAAGAGTAAGTTTGTGAAAAATTTCTAAATAGAAA 893  
 Db 6242 TTTTAAAGATAGATGGATGTAGAAATAGGTGTATGTATTAGAGTTTGTATAAAAAAG 6301  
 Qy 894 TGCATTTCGTCGAGCTAAAGCTTCTCTTTATTTTCCCTTTGTAGAAATGATTCGGT 953  
 Db 6302 TGTATTTTCGTTAGGTATAGTGGTTTATGTTTGTAAATTTTAGTATTTTGGAGGTTGAGT 6361  
 Qy 954 AGTAGGA 960  
 Db 6362 AGGAGGA 6368

RESULT 10  
 ABQ92787  
 ID ABQ92787 standard; DNA; 640681 BP.  
 XX AC ABQ92787;  
 XX DT 27-MAR-2002 (first entry)  
 XX DE Buchnera sp. genomic DNA SEQ ID NO:1.  
 XX KW Buchnera; cockroach-symbiotic bacterium; cockroach extermination;  
 KW circular; ds.  
 XX OS Buchnera sp.  
 XX PN JP2001292771-A.  
 XX PD 23-OCT-2001.  
 XX PF 07-APR-2000; 2000JP-0107160.  
 XX PR 07-APR-2000; 2000JP-0107160.  
 XX PA (RIKA ) RIKAGAKU KENKYUSHO.  
 XX DR WPI; 2002-126043/17.  
 XX A genomic DNA of cockroach-symbiotic bacterium -  
 PT Claim 1; Page 16-230; 237pp; Japanese.  
 XX The present invention describes a gene (I) derived from Buchnera sp.  
 CC containing the DNA (a) or (b), (a) has a fully defined base pair  
 CC sequence selected from a table of sequences found in the Buchnera sp.  
 CC genomic DNA of ABQ92787 given in the specification or is a DNA selected  
 CC from complementary DNA sequences, and (b) is a DNA which hybridises with  
 CC the DNA (a) and encodes a protein. Also described are: (1) a recombinant  
 CC vector (II) containing (I); (2) a transformant (III) containing (II);  
 CC (3) a genomic DNA of Buchnera sp. containing the sequence given in  
 CC ABQ92787; (4) a plasmid derived from Buchnera sp. containing DNA (c) or  
 CC (d) (c) is a DNA containing a fully defined sequence given in ABQ92788  
 CC or ABQ92789 and (d) is a plasmid which hybridises with a DNA; and (5) a  
 CC method for the preparation of a protein in which (III) is cultured and



DR WPI; 2001-602751/68.  
 XX Designing primers and probes for analysing diseases associated with  
 PT cytosine methylation state e.g. arthritis, cancer, aging,  
 PT arteriosclerosis comprising fragments of chemically modified genes  
 PT associated with cell cycle -  
 XX  
 XX Claim 1; SEQ ID NO 98; 28pp; English.  
 XX  
 XX Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA  
 CC molecules associated with the cell cycle and specific PCR primers of the  
 CC invention. The sequences are useful for detecting the methylation state  
 CC of all CpG dinucleotides in a sequence and therefore for analysing  
 CC associated diseases. By analysing cytosine methylations in the pretreated  
 CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy  
 CC of existing diseases or the predisposition to specific diseases can be  
 CC ascertained. The parameters may be compared to another set of genetic  
 CC and/or epigenetic parameters, the differences serving as basis for  
 CC diagnosis and/or prognosis events which are disadvantageous to patients.  
 CC The sequences of the invention are useful for the diagnosis and therapy  
 CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,  
 CC aging, glomerular disease, Lewy body disease, arthritis,  
 CC arteriosclerosis, solid tumours and cancers.  
 XX  
 XX Sequence 17419 BP; 5015 A; 311 C; 3864 G; 8229 T; 0 other;  
 SQ

Query Match 3.6%; Score 39.6; DB 22; Length 17419;  
 Best Local Similarity 49.5%; Pred. No. 3.7;  
 Matches 102; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

OY 852 GATCAACAAGAGTAAAGTTTGAAGAAATCTAAATAGAAATGCAATTCGTCGAGCTA 911  
 DB 2146 GTTTAGTAATATTAAGATGTTGAATGGTATAGAAAGAGTATATTTAGATTTTA 2205  
 OY 912 AAGCTTCCTCTCTTTATTTCTTTTCTAGAAATGATTCGGTAGGAAATATCGGTCG 971  
 DB 2206 AATTTATTTATTTTGTATTTTATTTATTTATTTTATTTTATTTTATTTTATTTT 2265  
 OY 972 AGTATGACCTTCGAGTATGCTGCTCTCGAAGACAGGAAAGTATGTTCTCAGCT 1031  
 DB 2266 ATTTATTTAGTGGAGTGTAGTGATGATTAATATAGTTTATTTGTAATTTTATTTTGGG 2325  
 OY 1032 TTAGAGCAATCTTCAGAAATATCTTT 1057  
 DB 2326 TTTAAGTTATTTTATTTTATTTAGTTT 2351

RESULT 13  
 ABL33295  
 ID ABL33295 standard; DNA; 17419 BP.  
 XX  
 XX ABL33295;  
 XX  
 XX 26-MAR-2002 (first entry)  
 XX  
 XX Human immune system associated gene SEQ ID NO: 1268.  
 DE  
 XX Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antianaemic; cytostatic; neotropic;  
 KW neuroprotective; anti-HIV; anticoagulant; ophthalmological;  
 KW antirheumatic; antiarthritis; antidyspeptic; antipsoriatic;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO200200928-A2.  
 PN  
 XX 03-JAN-2002.  
 PD  
 XX 02-JUL-2001; 2001WO-EP07537.  
 PF

XX  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 XX  
 XX Olek A, Piepenbrock C, Berlin K;  
 PI  
 XX WPI; 2002-130909/17.  
 DR  
 XX Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation -  
 XX  
 XX Claim 1; SEQ ID NO 1268; 32pp + Sequence Listing; German.  
 PS  
 XX The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 CC  
 XX Sequence 17419 BP; 5015 A; 311 C; 3864 G; 8229 T; 0 other;  
 SQ

Query Match 3.6%; Score 39.6; DB 24; Length 17419;  
 Best Local Similarity 49.5%; Pred. No. 3.7;  
 Matches 102; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

OY 852 GATCAACAAGAGTAAAGTTTGAAGAAATCTAAATAGAAATGCAATTCGTCGAGCTA 911  
 DB 2146 GTTTAGTAATATTAAGATGTTGAATGGTATAGAAAGAGTATATTTAGATTTTA 2205  
 OY 912 AAGCTTCCTCTCTTTATTTCTTTTGTAGAAATGATTCGGTAGGAAATATCGGTCG 971  
 DB 2206 AATTTATTTATTTTGTATTTTATTTATTTATTTTATTTTATTTTATTTTATTTT 2265  
 OY 972 ACTATGACCTTCGAGTATGCTGCTCTCGAAGACAGGAAAGTATGTTCTCAGCT 1031  
 DB 2266 ATTTATTTAGTGGAGTGTAGTGATGATTAATATAGTTTATTTGTAATTTTATTTTGGG 2325  
 OY 1032 TTAGAGCAATCTTCAGAAATATCTTT 1057  
 DB 2326 TTTAAGTTATTTTATTTTATTTAGTTT 2351

RESULT 14  
 ABK28238  
 ID ABK28238 standard; DNA; 17419 BP.  
 XX  
 XX ABK28238;  
 XX  
 XX 23-APR-2002 (first entry)  
 DT  
 XX DNA transcription associated complementary genomic DNA #56.  
 DE  
 XX DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;  
 KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;  
 KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;  
 KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;  
 KW immunological disorder; Werner syndrome; developmental disorder;  
 KW psoriasis; Kieffer's syndrome; neurological disorder; erythropoiesis;  
 KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;  
 KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;  
 KW angiodysplasia; congenital heart disease;  
 KW polyglutamine disorder; solid tumour.  
 XX  
 XX Unidentified.  
 OS  
 XX WO200192565-A2.  
 PN  
 XX

PD 06-DEC-2001.  
XX  
XX 06-APR-2001; 2001WO-EP03973.  
XX  
XX 06-APR-2000; 2000DE-1019058.  
XX  
XX 07-APR-2000; 2000DE-1019173.  
PR  
XX 30-JUN-2000; 2000DE-1032529.  
PR  
XX 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPIG-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI; 2002-090046/12.  
XX  
XX New nucleic acids or oligomers, useful for diagnosing or treating  
PT diseases associated with DNA transcription, e.g. immunological  
PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid  
PT tumours or cancer  
XX  
XX Claim 1; SEQ ID No 112; 32pp; English.  
XX  
XX The invention relates to a nucleic acid, which comprises a segment of the  
CC chemically pretreated DNA of genes associated with DNA transcription from  
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide  
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical  
CC to the chemically pretreated DNA of genes associated with DNA  
CC transcription. The set of oligomer probes are useful for detecting the  
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)  
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for  
CC diagnosing or treating diseases associated with DNA transcription  
CC (particularly with the methylation status), e.g. adenosine deaminase  
CC deficiency, viral infection, retroviral infection, Sezary syndrome,  
CC haematological disorders, immunological disorders, Werner syndrome,  
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,  
CC neurological disorders, neurodegenerative disorders, Waardenburg  
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial  
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart  
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours  
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription  
CC associated genomic DNA molecules of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format directly from the  
CC European Patent Office.  
XX  
XX Sequence 17419 BP: 5015 A; 311 C; 3864 G; 8229 T; 0 other;  
SQ

Query Match 3.6%; Score 39.6; DB 24; Length 17419;  
Best Local Similarity 49.5%; Pred. No. 3.7;  
Matches 102; Conservative 0; Mismatches 104; Indels 0; Gaps 0;  
QY 852 GATGACAAAGAGTAAGTCTTTGAAATCTCTAAATAGAAATGATCTTGTGTCGAGCTA 911  
DB 2146 GTTAGTAATATTAGAAATCTGTAATGGGTATAGAAAGAAATGATCTTAAAGATTTTTA 2205  
QY 912 AAGCTTCTCTCTCTTTATTCCTTTTGTAGATGATTCGGTAGTAGGAATATGGGGTCG 971  
DB 2206 AATTTATTTATTTTATTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 2265  
QY 972 AGTATGACGTGGAGTGAATGGCTGTTTCGGAGACAGGAAAGTCAATCTCAGCT 1031  
DB 2266 AATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 2325  
QY 1032 TTAGAGCAATCTTCAGATATACATTTT 1057  
DB 2326 TTTAAGTTATTTTATTTTATTTTATTTT 2351

RESULT 15  
AAF28550/C  
ID AAF28550 standard; DNA; 99629 BP.  
XX  
AC AAF28550;

XX 04-APR-2001 (first entry)  
DT  
XX Genomic fragment #37.  
DE  
XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;  
KW bronchopulmonary; endocarditis; meningitis; ss.  
KW  
XX Moraxella catarrhalis.  
OS  
XX WO200078968-A2.  
PN  
XX 28-DEC-2000.  
PD  
XX 16-JUN-2000; 2000WO-US16649.  
PF  
XX 18-JUN-1999; 99US-0140121.  
PR  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
XX Lagace RE, Patterson C, Berg KL;  
PI  
XX WPI; 2001-041427/05.  
XX  
XX Genomic library for identifying diagnostic and therapeutic  
PT compositions, and for identifying virulence factors, regulatory  
PT elements and drug targets, comprises Moraxella catarrhalis nucleic  
PT acids  
XX  
XX Claim 1; Page 391-415; 545pp; English.  
PS  
XX The present invention relates to a Moraxella catarrhalis genomic library  
CC comprising a combination of 41 nucleic acid molecules (see  
CC AAF28514-AAP28554). The library has a number of uses described in the  
CC specification e.g. is useful for identifying diagnostic and therapeutic  
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large  
CC aerobic gram-negative diplococcus, normally found among the bacterial  
CC flora of human upper airways. M. catarrhalis is known to cause acute,  
CC localised infections such as otitis media, sinusitis and bronchopulmonary  
CC infection and life-threatening, systemic diseases including endocarditis  
CC and meningitis.  
XX  
XX Sequence 99629 BP: 29233 A; 19222 C; 21909 G; 29264 T; 1 other;  
SQ

Query Match 3.6%; Score 39.6; DB 22; Length 99629;  
Best Local Similarity 61.8%; Pred. No. 7.2;  
Matches 63; Conservative 0; Mismatches 39; Indels 0; Gaps 0;  
QY 803 GCCATATCAGTGGAGCAACCTCCCTTTGATTCCTTATATGACAGAGATGACAAAG 862  
DB 8570 GCCATATTTGGCGATGGCAATTTACATTTAAATATTTCTAAACCTGCCCATTTGCCAAAG 8511  
QY 863 ACTAAGTTTTTGAAAAATCTAAAAATAGAAATGCAATTTGTGT 904  
DB 8510 ATGAATTTTTTGAAAAATGCCAAACGCTCAATACATTATGTGT 8469

RESULT 16  
ABL32333/C  
ID ABL32333 standard; DNA; 8895 BP.  
XX  
XX ABL32333;  
AC  
XX 26-MAR-2002 (first entry)  
DT  
XX Human immune system associated gene SEQ ID NO: 306.  
DE  
XX Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmologic;  
KW antineumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.  
XX Homo sapiens.  
OS  
PN WO200200928-A2.  
XX  
XX 03-JAN-2002.  
PD  
XX  
XX 02-JUL-2001; 2001WO-EP07537.  
PF  
XX  
XX 30-JUN-2000; 2000DE-1032529.  
PR  
XX 01-SEP-2000; 2000DE-1043826.  
PR  
XX (EPIC-) EPIGENOMICS AG.  
PA  
XX Olek A, Piepenbrock C, Berlin K;  
PI WPI; 2002-130909/17.  
XX  
XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation  
XX  
XX Claim 1; SEQ ID NO 306; 32pp + Sequence Listing; German.  
PS  
XX The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anemia, cancer, acute myeloid  
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
XX Sequence 8895 BP; 2269 A; 62 C; 1898 G; 4665 T; 1 other;  
SQ  
Query Match 3.6%; Score 39; DB 24; Length 8895;  
Best Local Similarity 46.2%; Pred. No. 4.2;  
Matches 129; Conservative 0; Mismatches 150; Indels 0; Gaps 0;  
QY 635 AATTCGAAGAATTTAATGACTGCTTGATGGTCTAAGAAATCAGAAAGGAGCCGAAAGCT 694  
DB 7126 AATACAAAAAATAAATACCACTTAATAAAAAAATATACATAAAAAAATAA 7067  
QY 695 TCTAAGTAATCAAGAAAAATCTTGGACAATAGCTTTTCCATAGAGAAACCGTCAGGAA 754  
DB 7066 AAAAAAACTAAAAAATAAATTAACAAACAATAAATAAATAAATAAATAAATAA 7007  
QY 755 AGGAGTCGATACCAATAATTCGGATTAAAGAGGGAGTGTACACAGCCATATCAGTG 814  
DB 7006 AAAAAAATAAATAATATCGAAAAACCTTAATACTACTACTAAACCTCTTACAT 6947  
QY 815 GAGCAACCTTCCCTTTGATCTTTGATTAATGACAGAGATGAACAAGAGTAAGTTTGG 874  
DB 6946 CTCGCCAAACCACTTTTACCCATCTATAAACAATAACCAATCTAATATATCTAA 6887  
QY 875 AAAAATCTAAATAGAAATGCAATTTGTGCGAGCTAAA 913  
DB 6886 AATAACCTAAAATCAAAACCAACCTCTTTCGTATACA 6848  
RESULT 17  
ABV61126/c  
ID ABV61126 standard; cDNA; 326 BP.  
XX  
XX AC ABV61126;  
XX  
XX 13-SEP-2002 (first entry)  
XX  
XX Human prostate expression marker cDNA 61117.  
XX  
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

KW pharmacogenomic marker; gene; ss.  
XX  
OS Homo sapiens.  
PN WO200160860-A2.  
XX  
XX 23-AUG-2001.  
PD  
XX  
XX 20-FEB-2001; 2001WO-US05171.  
PF  
XX  
XX 17-FEB-2000; 2000US-183319P.  
PR  
XX 16-MAR-2000; 2000US-189862P.  
PR  
XX 25-MAY-2000; 2000US-207454P.  
PR  
XX 09-JUN-2000; 2000US-211314P.  
PR  
XX 18-JUL-2000; 2000US-219007P.  
PR  
XX 13-DEC-2000; 2000US-255281P.  
PA  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
PI  
XX Schlegel R, Endege WO, Monahan JE;  
PI WPI; 2001-662795/76.  
DR  
XX  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer  
XX  
XX Claim 1; Page 11610; 11750pp; English.  
PS  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
XX Sequence 326 BP; 100 A; 48 C; 37 G; 140 T; 1 other;  
SQ  
Query Match 3.6%; Score 38.6; DB 23; Length 326;  
Best Local Similarity 49.0%; Pred. No. 1.6; 105; Indels 0; Gaps 0;  
Matches 101; Conservative 0; Mismatches 105; Indels 0; Gaps 0;  
QY 735 ATAGAGAAACCGTGAGGAAAGAGTCGATACCATATAATTCGGATTAAAAAGGAGTGC 794  
DB 268 AAAAACCAACCCCGGGGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 209  
QY 795 TTACAACAGCCATATCAGTGGAGCAACCTTCCCTTTGATTTCTTTGATAATGACAGAAGAT 854  
DB 208 TTACAAAAGGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 149  
QY 855 GAACAAGAGTAAGTTTGTGAAAAATTCCTAAATAAGAAATGCATTTCTGTCGAGCTAAAG 914  
DB 148 AACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 89  
QY 915 CTGCTTCTCTTTATTTTCTTTTGT 940  
DB 88 TTTTCTTTTCTTTTCTTTTCTTTT 63  
RESULT 18  
ABK33948  
ID ABK33948 standard; DNA; 18997 BP.  
XX  
XX AC ABK33948;



XX DT 18-JUN-2002 (first entry)

XX DE Human DNA for staging of Astrocytomas #16.

XX KW Human; ds; astrocytoma; cytostatic; staging; cysteine methylation; CpG; bisulphite; brain tissue; MALDI; ESI; electron spray mass spectrometry; matrix assisted laser desorption/ionization mass spectrometry.

XX KW OS Homo sapiens.

XX PN WO200202808-A2.

XX PD 10-JAN-2002.

XX PF 02-JUL-2001; 2001WO-EP07538.

XX PR 30-JUN-2000; 2000DE-1032529.

XX PR 01-SEP-2000; 2000DE-1043826.

XX PA (EPIC-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K;

XX DR WPI; 2002-171649/22.

XX PT Novel chemically modified genomic DNA sequences, useful in the treatment and/or diagnosis of astrocytomas or predisposition to astrocytomas.

XX PS Claim 1; SEQ ID NO 31; 37pp; English.

XX CC The invention relates to a nucleic acid comprising a sequence (I) of at least 18 bases in length of a segment of chemically pre-treated genomic DNA which has any one of the sequences of (ABK33919-ABK34032) or its complement. Also included are an oligonucleotide or peptide nucleic acid (or set thereof) of at least 9 nucleotides which hybridises to (I), primers for (I), probes for detecting cytosine methylation or single-nucleotide polymorphisms (SNP) in (I), an array of oligomers or peptide nucleic acids for analysing diseases associated with the methylation status of the CpG dinucleotides of (I). The array is useful for determining genetic and/or epigenetic parameters, classification, differentiation, grading, staging, treatment and/or diagnosis of astrocytomas, or the predisposition to astrocytomas by analysing cytosine methylations, involves obtaining a biological sample containing genomic DNA, extracting the genomic DNA, converting cytosine bases which are unmethylated at the 5-position, in the genomic DNA sample, to uracil or another base which is dissimilar to cytosine in terms of hybridisation behaviour, by chemical treatment and amplifying chemically pre-treated genomic DNA fragments using the array and a polymerase, where the amplification carries a detectable label. The method further involves identifying methylation status of one or more cytosine positions, and analysing methylation status of the cytosine positions by reference to one or more data sets. The genomic DNA is chemically treated by using a bisulphite, hydrogen sulphite or disulphite. The amplification step amplifies DNA which is of particular interest in astrocytoma or brain tissue, based on the specific genomic methylation status of brain tissues, as opposed to background DNA. The amplification carries a fluorescent label or radionuclide. Optionally, the labels of the amplification are detachable molecule fragments having a typical mass which are detected in a mass spectrometer. The fragments of chemically pre-treated genomic DNA to be amplified, have a single positive or negative charge for a better detectability in the mass spectrometer. Preferably, the amplification or fragments of the amplification are detected by matrix assisted laser desorption/ionization mass spectrometry (MALDI) or using electron spray mass spectrometry (ESI). The present sequence is one of the chemically pre-treated reference DNA samples of the invention.

XX CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 18997 BP; 4675 A; 299 C; 4016 G; 10007 T; 0 other;

XX Query Match 3.5%; Score 38.4; DB 24; Length 18997;

XX Best Local Similarity 54.2%; Pred. No. 8.2;

XX Matches 78; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

XX QY 856 AACAAAGAGTAAGTTTTCAAAATCTAAATAGAAATGCATTTGTCTCGAGCTAAAGC 915

XX DB 4017 AAAAAAATAAAATAGTTAAATTTTAAATATTTTATTTGTTAAATATATAAT 4076

XX QY 916 TTGCTTCTCTTAATTTTCCTTTTGTAGAATGATCGTAGTATGGGTGCGAGTA 975

XX DB 4077 TTGAATATATTTTAGTTTGTGAAAATGAGTTGAGAGAGTAGGTGAAATGAGTA 4136

XX QY 976 TGCACGTTGGAGTGATGGCTGTT 999

XX DB 4137 TGTTTTCGGGTTTTCGACGTT 4160

XX RESULT 19

XX ABL32570

XX ID ABL32570 standard; DNA; 18997 BP.

XX AC ABL32570;

XX DT 26-MAR-2002 (first entry)

XX DE Human immune system associated gene SEQ ID NO: 543.

XX KW Human; immune system disease; cytosine methylation; antiasthmatic; antiarteriosclerotic; antianaemic; cytostatic; nontropic; neuroprotective; anti-HIV; anticonvulsant; ophthalmological; antirheumatic; antiarthritic; antidiabetic; antipsoriatic; antinflammatory; cancer; eye disease; arteriosclerosis; anaemia; acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy; neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene; ds.

XX OS Homo sapiens.

XX PN WO200200928-A2.

XX PD 03-JAN-2002.

XX PF 02-JUL-2001; 2001WO-EP07537.

XX PR 30-JUN-2000; 2000DE-1032529.

XX PR 01-SEP-2000; 2000DE-1043826.

XX PA (EPIC-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K;

XX DR WPI; 2002-130909/17.

XX PT Nucleic acid comprising fragment of chemically modified gene, useful for diagnosis and treatment of diseases associated with abnormal cytosine methylation.

XX PS Claim 1; SEQ ID NO 543; 32pp + Sequence Listing; German.

XX CC The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders, including eye diseases such as retinopathy, neovascular glaucoma and macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel diseases. The present sequence is a gene of the invention.

XX SQ Sequence 18997 BP; 4675 A; 299 C; 4016 G; 10007 T; 0 other;



FT misc\_feature 25586 /\*tag= c  
FT FT /note= "This degenerate base represents polymorphic site  
FT (PS) 3"  
FT 26157  
FT misc\_feature /\*tag= d  
FT FT /note= "This degenerate base represents polymorphic site  
FT (PS) 4"  
XX WO200196350-A2.  
XX  
XX 20-DEC-2001.  
XX  
XX 18-JUN-2001; 2001WO-US19447.  
XX  
XX 16-JUN-2000; 2000US-212328P.  
XX  
XX (GENA-) GENAISSANCE PHARM INC.  
XX  
XX Finkel K, Koshy B, Tanguay DA;  
XX  
XX WPI; 2002-090198/12.  
XX  
XX New isolated polynucleotide, a polymorphic variant of glutamate  
XX receptor, metabotropic 3 (GRM3) gene for expressing GRM3 protein  
XX isoform to screen drugs to treat GRM3 activity-related disease  
XX  
XX Claim 20; Page 135-153; 165pp; English.  
XX  
XX The invention relates to an isolated polynucleotide which is a  
XX polymorphic variant of glutamate receptor, metabotropic 3 (GRM3)  
XX isogene. GRM3 is a receptor for glutamate, the major excitatory  
XX neurotransmitter in the mammalian central nervous system. Human GRM3  
XX located on chromosome 7q21.1-q21.2 is expressed in human foetal and adult  
XX whole brain especially in the caudate nucleus and corpus collosum. GRM3  
XX DNA is useful in gene therapy and also for studying the expression and  
XX function of GRM3. GRM3 polypeptide is used for screening drugs. A  
XX recombinant non-human organism is used to study expression of GRM3 SG  
XX in vivo, for in vivo screening and testing of drugs targetted against  
XX GRM3 protein, and for testing the efficacy of therapeutic agents and  
XX compounds for neurological disorders in a biological system. GRM3  
XX haplotypes are for treating diseases associated with GRM3 activity,  
XX e.g., neurological disorders. The present sequence is an allelic variant  
XX of human GRM3 gene fragment containing polymorphic sites (PS).  
XX  
XX Query Match 3.5%; Score 38.4; DB 24; Length 50000;  
XX Best Local Similarity 51.9%; Pred. No. 12;  
XX Matches 110; Conservative 0; Mismatches 101; Indels 1; Gaps 1;

QY 727 GCTTTCATAGAAAACCGTGGAGAAAGGAGTCCATACCATATAATCCGATTAAAAA 786  
DB 41412 GCCTTTCATAGATCAATCTTTATTAACACATCATTAACATTAATCAACAC 41353  
QY 787 GGAGTGCCTTACACAGCATATCAGTGGAGCAACCTTCCCTTTCATCTTGAATGA 846  
DB 41352 TAATGAGTTTACAAATACCTCTTATATATATTCAGATTTGGAAACAATTTTATT 41293  
QY 847 CAGAGATGACAAAGTAAGTATTTTGAAGAAATTTCTAAATAGAAATGCAATTCGTCG 906  
DB 41292 CTGTGAATGCATAG-GTAAAGCTGAGCCAAATGACAAATGCAATTCCTTTCTAA 41234  
QY 907 AGCTPAAAGCTGCTCTCTCTTTATTTCTTTT 938  
DB 41233 ATACAAACCTTGTGTTTTTTTCTTTCTTTT 41202

RESULT 22  
ABQ67094  
ID ABQ67094 standard; DNA; 83391 BP.  
XX  
XX ABQ67094;  
AC

XX 28-AUG-2002 (first entry)  
XX  
XX Human angiogenesis associated polynucleotide SEQ ID NO 124.  
XX  
XX Human: angiogenesis; methylation; eye disease; glaucoma; tumour;  
XX inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcerds;  
XX macular degeneration; inflammatory bowel disease; Crohn's disease;  
XX antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
XX antiarteriosclerotic; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200246454-A2.  
XX  
XX 13-JUN-2002.  
XX  
XX 06-DEC-2001; 2001WO-EP14320.  
XX  
XX 06-DEC-2000; 2000DE-1061338.  
XX  
XX (EPIG-) EPIGENOMICS AG.  
XX  
XX Schacht O;  
XX  
XX WPI; 2002-500450/53.  
XX  
XX New nucleic acid fragments from chemically treated  
XX angiogenesis-associated genes, useful for determining methylation  
XX status, e.g. in diagnosis or treatment of cancer  
XX  
XX Claim 1; SEQ ID NO 124; 41pp + Sequence Listing; German.  
XX  
XX The invention relates to a nucleic acid (I) comprising a segment of 18  
XX bases of chemically pretreated DNA of angiogenesis-associated genes (II)  
XX having sequences (ABQ66971-ABQ67178) or their complements (I), also  
XX related oligomers, are used to evaluate the methylation status and/or  
XX single-nucleotide polymorphisms in angiogenesis-related genes, for  
XX diagnosis and treatment of eye diseases, proliferative retinopathy,  
XX neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,  
XX diabetic retinopathy, macular degeneration caused by neovascularisation,  
XX psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and  
XX Crohn's disease.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 83391 BP; 24547 A; 665 C; 16953 G; 41209 T; 17 other;  
XX  
XX Query Match 3.5%; Score 38.4; DB 24; Length 83391;  
XX Best Local Similarity 52.5%; Pred. No. 14;  
XX Matches 84; Conservative 0; Mismatches 76; Indels 0; Gaps 0;  
QY 344 TTGCTTTGTAGGGTTTGTATGCGTTGTTTTTTACTAGGAAGAGCTTGTCTTTTCTATG 403  
DB 80412 TTTATTTAGAGCGGTTGTATAAAGAGGTTTTTTTGTGGAAGGAATGTTTTTTTATTT 80471  
QY 404 ACTTTAGTGAAGAGTCTCCCATGTTTATGAAAAATTCAGAAAAATGAGGGATGTAGAAGC 463  
DB 80472 GGATTTTGTGTTGAGGTTTTTATTTATTTAAAGAGTTGATAGAGAGGGAGGAAAAGG 80531  
QY 464 CAAGGCACGTGCCCTFAAAGGTCTATGATCAGATCAGAGAG 503  
DB 80532 AGAGGAAGGGGATCGGAGGTTGGGGGAGGGGGAGGG 80571  
XX  
XX RESULT 23  
XX AAS46333  
XX ID AAS46333 standard; DNA; 7667 BP.  
XX  
XX AAS46333;  
XX  
XX 18-DEC-2001 (first entry)  
XX  
XX

XX Tumour suppressor gene derived chemically modified sequence #55.  
DE Human; tumour suppressor gene; oncogene; antitumour; cytostatic;  
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;  
KW cytosine methylation; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200168912-A2.  
XX  
PD 20-SEP-2001.  
XX  
PF 15-MAR-2001; 2001WO-EP02955.  
XX  
PR 15-MAR-2000; 2000DE-1013847.  
PR 06-APR-2000; 2000DE-1019058.  
PR 07-APR-2000; 2000DE-1019173.  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
DR WPI; 2001-602752/68.  
XX  
XX Fragments of chemically modified genes associated with tumour suppressor  
PT genes and oncogenes, useful in designing primers and probes for  
PT analysing diseases associated with cytosine methylation state e.g.  
PT cancer  
XX  
PS Claim 1; SEQ ID No 55; 27pp; English.  
XX  
CC The invention relates to a nucleic acid comprising a sequence of 18  
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with  
CC bisphite, of genes associated with tumour suppression and  
CC oncogenes having a sequence taken from 536 (actually 533 since  
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences  
CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a  
CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may  
CC form part of a set of probes for detecting the cytosine methylation state  
CC and/or single nucleotide polymorphisms and also to be used in an  
CC array for analysing diseases associated with CpG dinucleotides e.g.  
CC cancers and tumours. The probes can also be used in a method for  
CC ascertaining genetic and/or epigenetic parameters for the diagnosis  
CC and/or therapy of existing diseases or the predisposition to specific  
CC diseases, by analysing cytosine methylations. The parameters may be  
CC compared to another set of genetic and/or epigenetic parameters, the  
CC differences serving as basis for diagnosis and/or prognosis events which  
CC are disadvantageous to patients. The present sequence is one of the  
CC 533 genomic sequences derived from tumour suppressor genes and  
CC oncogenes.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from Wipo at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 7667 BP; 2446 A; 63 C; 1493 G; 3665 T; 0 other;  
Query Match 3.5%; Score 38.2; DB 22; Length 7667;  
Best Local Similarity 49.7%; Pred. No. 6.6;  
Matches 97; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 566 TGCAACAGACAGCAAGATTGATGGGAGCTACAAATCCTGAAACGGGATAATAGGAAAAA 625  
DB 2674 TGTATTATAAAAAATTTTATTATTAGATTTTAAATTTTGTGGAGAAAAATAAGATAAA 2733  
QY 626 TGTCTTTCTTAATTCGAAGAAATTAATGACTGCTTGATGGTCTAAGAAATCAAGAGGAGA 685  
DB 2734 AATATGATAGTTAATGATTTTAAATAATAATAATTTTAAAGATAGNAGATATGAGG 2793  
QY 686 CCGAAGCTTCTAAGTAATCAGAAAAAATCTTGGCAATAGCTTTTCCATAGAGAAAC 745

DB 2794 TAGAAGTAATATTATTATTAAAGGAAATGTAGAGAAAAATATTGTTATTAGATAAAGGAG 2853  
QY 746 CGTGAGGAAAGGAGCT 760  
DB 2854 AGAGAGAGAGAGATT 2868  
RESULT 24  
ABU32294/G  
ID ABL32294 standard; DNA; 9415 BP.  
XX  
AC ABL32294;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Human immune system associated gene SEQ ID NO: 267.  
XX  
KW Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200200928-A2.  
XX  
PD 03-JAN-2002.  
XX  
PF 02-JUL-2001; 2001WO-EP07537.  
XX  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
DR WPI; 2002-130909/17.  
XX  
PT Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation  
XX  
PS Claim 1; SEQ ID NO 267; 32pp + Sequence Listing; German.  
XX  
CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
SQ Sequence 9415 BP; 2021 A; 297 C; 2877 G; 4220 T; 0 other;  
Query Match 3.5%; Score 38.2; DB 24; Length 9415;  
Best Local Similarity 48.8%; Pred. No. 7.1;  
Matches 103; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 595 TACAATCTCTGAAACGGGATAATAGGAAAAATGTGTTCTTAATTCGAAGAAATTAATGAC 654  
DB 4049 TAAACCTCCCCACACCTATTACACGAATAAATTTTCTTAATCTACCAAAATAAAAT 3990  
QY 655 TGCCTTGATGGTCTAAGAAATACAGAGAGACCGAAAGCTTCTAAGTAATCAAGAAAAA 714  
DB 3989 TCCCTTCCCACTTCCCTACCAAAAAATTTATCCAAACAATACCAAAACATCCGTAAAAAC 3930

QY 715 TCTTGGACATAGCTTTTCCATAGAGAAAACCGTGGAGAAAGGAGTCGATACCATTAAT 774  
Db 3929 TAAACAAAATAAACTATTTTAAATAAACCTAAAAAATAAAATCAATTAATCAAAAT 3870  
QY 775 CGATTAAAAAGGAGTCTTACACAGCC 805  
Db 3869 AAAATAAAAAATAATTAATAAATCC 3839

RESULT 25  
ABL32467/c  
ID ABL32467 standard; DNA; 15373 BP.  
XX  
AC ABL32467;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Human immune system associated gene SEQ ID NO: 440.  
XX  
KW Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antihaemic; cytostatic; neotropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antineumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200200928-A2.  
XX  
XX 03-JAN-2002.  
XX  
XX 02-JUL-2001; 2001WO-EP07537.  
XX  
XX 30-JUN-2000; 2000DE-1032529.  
XX  
XX 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPIG-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI; 2002-130909/17.  
XX  
XX Nucleic acid comprising fragment of chemically modified gene, useful  
XX for diagnosis and treatment of diseases associated with abnormal  
XX cytosine methylation  
XX  
XX Claim 1; SEQ ID NO 440; 32pp + Sequence Listing; German.  
XX  
XX The present invention provides a number of human immune system associated  
XX genes which are modified by the methylation of cytosines. The sequences  
XX can be used in the diagnosis and treatment of immune system disorders,  
XX including eye diseases such as retinopathy, neovascular glaucoma and  
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
XX diseases. The present sequence is a gene of the invention.  
XX  
XX Sequence 15373 BP; 5097 A; 89 C; 2465 G; 7722 T; 0 other;  
XX

Query Match 3.5%; Score 38.2; DB 24; Length 15373;  
Best Local Similarity 44.9%; Pred. No. 8.6;  
Matches 145; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

QY 606 AAGCGGATAATAGAAAATGTGTTCTAATTCGAAGAAATTAATGACTGCTTGATGGT 665  
Db 10266 AACTTTATACAAAAAARAACTAACTACTTCGCTCTACTTCATCCACTATTATTT 10207  
QY 666 CTAGAAATCAGAAAGGACCGAAGCTCTAAGTAATCAAGAAAAATCTTGACAAAT 725  
Db 10206 AACACAAAAAATAATAAATACTAATTTCAAAAAACTTTAATAATTTATCTTT 10147

QY 726 AGCTTTTCCATAGAGAAAACCGTGGAGAAAGGAGTCGATACCATTAATTCGATTAATA 785  
Db 10146 CTCTCTCACAAAAAATAATAATAAAAAAATAATCAATTAATTCAAAAATAATTC 10087  
QY 786 AGGAGTCTTACACAGCCATATCAGTGGAGCAACTCCCTTTGATTCTTTTGATAATG 845  
Db 10086 CTCTTAACCTCCAAATATATAAAGCTTAACCTAACCTAACCTAACCTAACCTAACCT 10027  
QY 846 ACAGAGATGAACAAGAGTAAGTTTGAATAATCTTAATAATAGAAATGCATTTGTGTC 905  
Db 10026 TAAAAAATAATCGCTTATATATATTTTAAAAAATCTATATAATAATAATAATTTCTT 9967  
QY 906 GAGCTAAAGCTTCTCTCTTTA 928  
Db 9966 ATACTAACTTATTTTCATATTA 9944

RESULT 26  
ABK31510/c  
ID ABK31510 standard; DNA; 47108 BP.  
XX  
AC ABK31510;  
XX  
DT 23-APR-2002 (first entry)  
XX  
DE Signal transduction associated gene modified DNA #177.  
XX  
KW Human; signal transduction associated gene; cytosine methylation state;  
KW CpG island; signal transduction associated disease; solid tumour; cancer;  
KW antitumour; cytostatic; mutant; ds.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200200926-A2.  
XX  
XX 03-JAN-2002.  
XX  
XX 29-JUN-2001; 2001WO-EP07472.  
XX  
XX 30-JUN-2000; 2000DE-1032529.  
XX  
XX 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPIG-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI; 2002-147896/19.  
XX  
XX Oligonucleotide for diagnosis and therapy of diseases associated with  
XX signal transduction e.g. cancer, comprises chemically modified genomic  
XX sequences of genes associated with signal transduction  
XX  
XX Claim 1; SEQ ID No 353; 24pp; English.  
XX  
XX The present invention relates to chemically modified DNA sequences of  
XX signal transduction associated genes. The DNA sequences are chemically  
XX modified using a solution of bisulphite, hydrogen sulphite or  
XX disulphite. Also disclosed are oligonucleotides and/or PNA oligomers  
XX for detecting the cytosine methylation state (CpG islands) of these  
XX genes, and a method for the diagnosis and/or therapy of genetic and  
XX epigenetic parameters of genes associated with signal transduction which  
XX The genomic DNA can be obtained from cells or cellular components which  
XX contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,  
XX cerebral-spinal fluid, tissue embedded in paraffin such as tissue from  
XX eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver,  
XX histologic object slides, and all their possible combinations. The  
XX sequences of the invention are useful for the diagnosis and therapy of  
XX diseases associated with signal transduction e.g. solid tumours and  
XX cancer. ABK31158-ABK31545 represent chemically pretreated genomic DNA  
XX sequences of different genes associated with signal transduction, or  
XX their complementary sequences.



including eye diseases such as retinopathy, neovascular glaucoma and macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel diseases. The present sequence is a gene of the invention.

PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138840.  
PR 16-JUN-1999; 99US-0139111.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140354.  
PR 28-JUN-1999; 99US-0140699.  
PR 29-JUN-1999; 99US-0140822.  
PR 30-JUN-1999; 99US-0140991.  
PR 01-JUL-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144333.  
PR 20-JUL-1999; 99US-0144333.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147933.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 23-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151203.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 23-SEP-1999; 99US-0155659.  
PR 24-SEP-1999; 99US-0156458.  
PR 28-SEP-1999; 99US-0156596.  
PR 29-SEP-1999; 99US-0157117.  
PR 04-OCT-1999; 99US-0157753.  
PR 05-OCT-1999; 99US-0157865.  
PR 06-OCT-1999; 99US-0158029.  
PR 07-OCT-1999; 99US-0158232.  
PR 08-OCT-1999; 99US-0158369.  
PR 12-OCT-1999; 99US-0158232.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.



PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
 Query Match            3.4%; Score 37.4; DB 21; Length 870; Best Local Similarity 51.5%; Pred.No. 4.8; Matches 86; Conservative 0; Mismatches 81; Indels 0; Gaps 0;		
OY	30 ATTTCTCTTGATGTAAAGGGGAATTGATTCTAACAGACCTTGGTCGAGCATCCATTCFA	89
DB	168 ATTTCTCCGGTGGTAGGCGGCCGTACCAGAATCTAGAGGAGCGAGTCATTTTG	109
OY	90 ATTTTGAAGTGAAGTCTACCAAGTTCATCAAATGTCTTATTGAAAAATAAAGTCTCTTG	149
DB	108 TTTTGGTTTTGAGTCTCTGTTTTAGGGGAAGAGGAAATGTGAAAGACAGAGAAGAGAG	49
OY	150 GAGTGTGATGTATTGAAGGGAAGACAATAATCTGATCCAATTTCTAAA	196
DB	48 ACTTTGTTGTTGTTTAGAGGAGAAGAATCTGATCCAAGTTATAA	2
 RESULT 31 ABL32359 ID ABL32359 standard; DNA; 6237 BP. XX AC ABL32359; XX XX DT 26-MAR-2002 (first entry) DE Human immune system associated gene SEQ ID NO: 332. XX DE XX Human; immune system disease; cytosine methylation; antiasthmatic; KW antiarteriosclerotic; antitanaemic; cytostatic; neurotropic; KW neuroprotective; anti-HIV; anticoagulant; ophthalmological; KW antirheumatic; antiarthritic; antididiabetic; antipsoriatic; KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia; KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy; KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; KW gene; ds. XX KW XX Homo sapiens. OS WO200200928-A2. PN XX PD 03-JAN-2002. XX PD PF 02-JUL-2001; 2001WO-EPO7537. XX PF PR 30-JUN-2000; 2000DE-1032529. PR 01-SEP-2000; 2000DE-1043826. XX PR PA (EPTG-) EPIGENOMICS AG. XX PA XX Olek A, Piepenbrock C, Berlin K; PI XX WPI; 2002-130909/17. XX DX XX Nucleic acid comprising fragment of chemically modified gene, useful PT for diagnosis and treatment of diseases associated with abnormal PT cytosine methylation - XX PT XX Claim 1; SEQ ID NO 332; 32pp + Sequence Listing; German. XX PS XX The present invention provides a number of human immune system associated CC genes which are modified by the methylation of cytosines. The sequences CC can be used in the diagnosis and treatment of immune system disorders, CC including eye diseases such as retinopathy, neovascular glaucoma and CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel CC diseases. The present sequence is a gene of the invention.		



```
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 13-AUG-1999; 99US-0149368.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150984.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159304.
PR 13-OCT-1999; 99US-0159305.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.4%; Score 37.2; DB 21; Length 522;
Best Local Similarity 51.9%; Pred No 4.5; 78; Indels 0; Gaps 0;
Matches 84; Conservative 0; Mismatches 0;

QY 30 ATTTCTCTGATTGTAAGGGGAATTGATTCACAGACCTTGGTCGAGCATCCATTCTA 89
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 168 ATTTCTTCCGGTGGGAGGCGACCGCGTACCAAGGATCTAGAGGCGGAGTCAATTTG 109
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 90 ATTTGAAGTGAGTCTACGAGTTCATCAAAATGCTTTATTGAAATTAAGTCTCTTTGTG 149
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 108 TTTTGGTTTGGATCTCTGTTTATAGGGAAGAGAAATGTGAAGACAGAGGAAGAGAG 49
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 150 GAGTTGATGTATTGAAGGGAAGACGAATAATCTGATCCAAATT 191
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 48 ACTTTGTTGTTGTTTATAGGGAAGAGAAATCTGATCCAAAGT 7
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 34
AAC51121/c
ID AAC51121 standard; DNA; 869 BP.
XX
AC AAC51121;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 67351.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
protein identification; signal transduction pathway;
metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 25-MAY-1999; 99US-0136021.
```

PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139453.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 24-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142058.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144003.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 06-OCT-1999; 99US-0157753.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 13-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.





PS Claim 1; SEQ ID NO 1693; 32pp + Sequence Listing; German.  
 XX The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and,  
 CC macular degeneration, arteriosclerosis, anemia, cancer, acute myeloid  
 CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 XX  
 SQ Sequence 7190 BP; 1434 A; 310 C; 2114 G; 3261 T; 71 other;  
 Query Match 3.4%; Score 37.2; DB 24; Length 7190;  
 Best Local Similarity 49.5%; Pred. No. 12;  
 Matches 90; Conservative 0; Mismatches 92; Indels 0; Gaps 0;  
 QY 837 TTGATAATGACAGAGATGAACAAAGAGTAAGTTTTCGAAAATCTTAAATAGAAATGC 896  
 DB 6770 TAGATGANTAGTTATGAAGATAATATTGNATGTTTINAGAAGAAATGTNAGATTATTTT 6829  
 QY 897 ATTGTGTCGAGCTAAAGCTTGCTCTCTTTATTTTCCTTTTGTAGAAATGATTCGGTAGT 956  
 DB 6830 GGAAGTGTGGGGAAGGTGGAAGTGATATTATGATTTGTTAGTAGAAAGTTACGTTAGA 6889  
 QY 957 AGGAATATGGGTCGAGTATGACCTTGGAGTGATTTGCTGCTGCAAGAACAGGGA 1016  
 DB 6890 GTTTATATAGTTTGTATTTTGTATAGAGTGGGAGGGGTGTAGGGGAAGAGAGGTAA 6949  
 QY 1017 GT 1018  
 DB 6950 TT 6951  
 RESULT 39  
 ABV52129/c  
 ID ABV52129 standard; cdna; 619 BP.  
 XX  
 AC ABV52129;  
 XX  
 DT 17-SEP-2002 (first entry)  
 XX  
 DE Human prostate expression marker cdna 52120.  
 XX  
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200160860-A2.  
 XX  
 PD 23-AUG-2001.  
 XX  
 PF 20-FEB-2001; 2001WO-US051171.  
 XX  
 PR 17-FEB-2000; 2000US-183319P.  
 PR 16-MAR-2000; 2000US-189862P.  
 PR 25-MAY-2000; 2000US-207454P.  
 PR 09-JUN-2000; 2000US-211314P.  
 PR 18-JUL-2000; 2000US-219007P.  
 PR 13-DEC-2000; 2000US-255281P.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX  
 PI Schlegel R, Endege WO, Monahan JE;  
 XX  
 DR WPI; 2001-662795/76.  
 XX  
 XX Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer -  
 PT

PS Claim 1; Page 10116; 11750pp; English.  
 XX The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement. (I) is useful for:  
 CC (a) assessing whether a patient is afflicted with prostate cancer;  
 CC (b) monitoring the progression of prostate cancer in a patient;  
 CC (c) assessing the efficacy of a test compound to inhibit prostate  
 CC cancer in a patient;  
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
 CC in a patient; a composition for inhibiting prostate cancer in a patient;  
 CC (e) selecting the prostate cell carcinogenic potential of a compound;  
 CC (f) determining whether prostate cancer has metastasized in a patient;  
 CC (g) assessing the aggressiveness or indolence of prostate cancer in a  
 CC patient;  
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
 CC patient;  
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
 XX  
 SQ Sequence 619 BP; 265 A; 95 C; 107 G; 149 T; 3 other;  
 Query Match 3.4%; Score 37; DB 23; Length 619;  
 Best Local Similarity 60.4%; Pred. No. 5.4;  
 Matches 61; Conservative 0; Mismatches 40; Indels 0; Gaps 0;  
 QY 330 CAATAGGCGATCAGATTGCTTTGTAGGGTTTGTATCGGTGTTTCTTACTAGGAAGAGCTT 389  
 DB 427 CATAAAGCGCTTAGAAGCTTTCTTAAAGTTTCTTCTGTTGTTTCTTCTCCCATAGATT 368  
 QY 390 GTACTTTTCTTATGACTTTAGTGAGAGTCTTCCCATGTTT 430  
 DB 367 TTTGTTTTCATATTTGTAGTCAAGCTTTGCTTATGTTT 327  
 RESULT 40  
 AA231005  
 ID AA231005 standard; DNA; 1134 BP.  
 XX  
 AC AA231005;  
 XX  
 DT 05-JAN-2000 (first entry)  
 XX  
 DE Partial dnaN gene.  
 XX  
 KW Gram positive bacteria; dnaE; dnaX; dnaB; PolC; dnaN; dnaG; helicase;  
 KW alpha subunit; DNA polymerase III holoenzyme; gamma subunit; tau subunit;  
 KW clamp loader; glue protein; replication; antibiotic; ss.  
 XX  
 OS Staphylococcus aureus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1134  
 FT /\*tag= a  
 FT /note= "Dna N gene product"  
 XX  
 PN W09937661-A1.  
 XX  
 PD 29-JUL-1999.  
 XX  
 PF 25-JAN-1999; 99WO-US01547.  
 XX  
 PR 27-JAN-1998; 98US-0074522.  
 PR 22-JUL-1998; 98US-0093727.  
 XX  
 XX (UYRQ ) UNIV ROCKEFELLER;  
 XX  
 XX O'Donnell ME, Zhang D, Whipple R;  
 XX  
 DR WPI; 1999-590685/50.  
 DR P-PSDB; AAY49071.  
 XX  
 XX New isolated dnaE, dnaX and dnaB genes from Gram positive bacteria,  
 PT used to develop screening assays for identifying antibiotic compounds -  
 PT

XX PS Example 8; Page 30; 132pp; English.

XX CC This sequence is the partial dnaN gene of Staphylococcus aureus. The

XX CC invention relates to a number of isolated DNA molecules from Gram

XX CC positive bacterium, corresponding to dnaE (AAZ31001), dnaX

XX CC (AAZ31002), and dnaB (AAZ31003). The polC, dnaE, and dnaB genes

XX CC (AAZ31004-231006) are also identified. The dnaE gene corresponds to the

XX CC alpha subunit of the Escherichia coli, DNA polymerase III holoenzyme,

XX CC dnaX corresponds to the gamma and tau subunits, and dnaB corresponds to

XX CC the helicase. The alpha subunit is the actual DNA polymerase, the gamma

XX CC complex forms the clamp loader and tau is a "glue protein". DnaX encodes

XX CC both gamma and tau. Tau is the product of the full gene, while gamma is

XX CC the product of the first two thirds of the gene. DnaN forms the beta

XX CC subunit which forms the sliding clamp, and dnaG encodes a primase. The

XX CC DNA sequences of the invention can be used to identify agents that

XX CC inhibit or promote DNA replication by acting on various parts of the gram

XX CC positive bacterial DNA polymerase holoenzyme. The products and methods of

XX CC the invention can be used for identifying pharmacological agents or lead

XX CC compounds for agents active at the level of a replication protein

XX CC function, particularly DNA replication. The agents identified can be used

XX CC as antibiotics.

XX SQ Sequence 1134 BP; 411 A; 176 C; 194 G; 353 T; 0 other;

Query Match 3.4%; Score 37; DB 20; Length 1134;

Best Local Similarity 50.3%; Pred. No. 6.8;

Matches 91; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 34 CTCCTGATTTGTAAGGGGAATTGATCTTAACAGACCTTGGTCGAGCATCCATCTCAATTT 93

DB 834 CTCCTTATTAGCAGCTGAAGGTGGTAATACCGTTATTAAATTAAGTACAGGTGATGACGT 893

QY 94 TGAAGTGAGTTCTACAGTTCACAAAATGCTTATTGAAAATAAAGTCTCTTGTGGAGT 153

DB 894 TGTGTAATATCTCTTACATCACCAAGAAATGGTACTGTAAGAAAGAAAGTGTGATGCAAA 953

QY 154 TGATGTTATGAAGGGAAGACGAATAATCTGATCCAAATCTTAAAGGAGACACGATCGTT 213

DB 954 CGATGTTGAAGGTGGTGCTGAAATTTTCATTCAACTCTAAATATATATGATGATGCTTT 1013

QY 214 A 214

DB 1014 A 1014

RESULT 41

AAF54735

ID AAF54735 standard; DNA; 1134 BP.

XX AC AAF54735;

XX DT 15-MAY-2001 (first entry)

XX DE Nucleotide sequence of a dnaN polypeptide.

XX KW dnaE; Gram positive bacteria; polC; dnaE; hola; holB; dnaX; dnaN; ssb;

XX KW dnaG; antibiotic; replication; cell growth; cell death;

XX KW bacterial infection; ss.

XX OS Staphylococcus aureus.

XX FT Key Location/Qualifiers

XX FT CDS 1..1134

XX FT /tag= a

XX FT /product= "dnaN protein"

XX PN W0200109164-A2.

XX PD 08-FEB-2001.

XX PF 28-JUL-2000; 2000WO-US20666.

XX XX

PR 29-JUL-1999; 99US-0146178.

XX (UVRQ ) UNIV ROCKEFELLER.

XX O'Donnell ME, Bruck I, Zhang D, Whipple R;

XX WPI; 2001-147453/15.

XX P-PSDB; AAB31935.

XX Isolated DNA molecule from a Gram positive bacterium encoding DNA

XX replication proteins used to identify compounds which have antibiotic

XX activity -

XX Disclosure; Page 39-40; 239pp; English.

XX The present sequence encodes a dnaN polypeptide. The specification

XX describes DNA molecules from Gram positive bacteria, which comprise

XX a coding region from a polC, dnaE, hola, holB, dnaX, ssb, dnaG

XX or a dnaB gene. These sequences encode proteins that replicate the

XX chromosome of Gram positive bacteria. They are used for sequencing

XX and amplification of DNA and in drug discovery to identify compounds

XX which have antibiotic activity through interference with replication.

XX They are used in methods for identifying compounds that are active at

XX the level of DNA replication and result in arrest of cell growth or

XX cell death of bacteria to treat bacterial infections in animals.

XX SQ Sequence 1134 BP; 411 A; 176 C; 194 G; 353 T; 0 other;

Query Match 3.4%; Score 37; DB 22; Length 1134;

Best Local Similarity 50.3%; Pred. No. 6.8;

Matches 91; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 34 CTCCTGATTTGTAAGGGGAATTGATCTTAACAGACCTTGGTCGAGCATCCATCTCAATTT 93

DB 834 CTCCTTATTAGCAGCTGAAGGTGGTAATACCGTTATTAAATTAAGTACAGGTGATGACGT 893

QY 94 TGAAGTGAGTTCTACAGTTCACAAAATGCTTATTGAAAATAAAGTCTCTTGTGGAGT 153

DB 894 TGTGTAATATCTCTTACATCACCAAGAAATGGTACTGTAAGAAAGAAAGTGTGATGCAAA 953

QY 154 TGATGTTATGAAGGGAAGACGAATAATCTGATCCAAATCTTAAAGGAGACACGATCGTT 213

DB 954 CGATGTTGAAGGTGGTGCTGAAATTTTCATTCAACTCTAAATATATATGATGATGCTTT 1013

QY 214 A 214

DB 1014 A 1014

RESULT 42

AAH32611/C

ID AAH32611 standard; cDNA; 1075 BP.

XX AC AAH32611;

XX DT 10-AUG-2001 (first entry)

XX DE Human secreted protein gene 27 cDNA clone HJAAJ58, SEQ ID NO:100.

XX KW Human; secreted protein; proliferative disorder; cancer;

XX KW fetal abnormality; developmental abnormality; haematopoietic disorder;

XX KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;

XX KW inflammation; allergy; neurological disorder; Alzheimer's disease;

XX KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;

XX KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;

XX KW cardiovascular disorder; angioinetic disorder; kidney disorder;

XX KW gastrointestinal disorder; pregnancy-related disorder; tumour;

XX KW endocrine disorder; infection; wound healing; vulnery;

XX KW cell culture; chemotaxis; food additive;

XX KW binding partner identification; ss.

XX OS Homo sapiens.

XX XX





CC alleviating symptoms associated with the disorders mentioned above, and  
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
CC immunosorbent assay (ELISA). The present sequence represents a human  
CC secreted protein-encoding cDNA of the invention.

XX Sequence 1724 BP; 433 A; 242 C; 279 G; 770 T; 0 other;

Query Match 3.4%; Score 36.8; DB 22; Length 1724;  
Best Local Similarity 52.6%; Pred. No. 9;  
Matches 103; Conservative 0; Mismatches 92; Indels 1; Gaps 1;

Qy 694 TTCTAAGTATCAAGAAAAATCTTTGGCAATAGCTTTTCCATAGAGAAAAACCGTGAGGA 753

Db 1069 TTGTTTGTAAAAAATAAACAATAAATAAATGTTAGCTAAACAAGTAAGAAAAAAG 1010

Qy 754 AAGGAGTCGATACCATATAATCCGATTTAAAAAGGGAGTGCTTACACAGCCATATCACT 813

Db 1009 AAGACTCAATTAATGAATCAGAAATGAAAAGGAGACATGCAATTCATACACACAGA 950

Qy 814 GGAGCACTTCCTCTTGATCTTTGATAATGACAGAAGTGAACAAGAGTAAGTTT 872

Db 949 AATCAAGAATCATTTGGAGTCTATTACATCAACCATGATGATCAACAAATGGAAGCCT 890

Qy 873 TGAATAATCTTAAT 888

Db 889 AGTAAAAATGTATAAT 874

#### RESULT 44

ID AAK68427 standard; DNA; 1761 BP.

XX AAK68427;

DT 06-NOV-2001 (first entry)

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23239.

KW Human; Immune; haematopoietic; immune/haematopoietic antigen; cancer;

XX cytostatic; gene therapy; vaccine; metastasis; ds.

OS Homo sapiens.

XX WO200157182-A2.

PD 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216680.

XX 11-JUL-2000; 2000US-0217487.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 14-AUG-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231142.  
PR 08-SEP-2000; 2000US-0231143.  
PR 08-SEP-2000; 2000US-0231144.  
PR 08-SEP-2000; 2000US-0231145.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234597.  
PR 25-SEP-2000; 2000US-0234598.  
PR 26-SEP-2000; 2000US-0234584.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 02-OCT-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239335.  
PR 13-OCT-2000; 2000US-0239337.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241321.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0241826.  
PR 08-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.





Db 759 AAAGACTCAATAAATGAATCAGACATCAAAAAGGAGACATGACAAATGATACCACAGA 818  
QY 814 GGAGCAACCTTCCTTTGATTTGATAATGACAGAGA-TGAACAAAGAGTAAGTTTT 872  
Db 819 AATACAAGAATCATTTGGAGTCTATTACAATCAACCATGAGTCAACAAATTTGAAAGCCT 878  
QY 873 TGAATAAATTTCTAAAT 888  
Db 879 AGTAAATATGATAAT 894

RESULT 46  
ABQ67075  
ID ABQ67075 standard; DNA: 7857 BP.  
XX  
AC ABQ67075;  
XX  
DT 28-AUG-2002 (first entry)  
XX  
DE Human angiogenesis associated polynucleotide SEQ ID NO 105.  
XX  
KW Human; angiogenesis; methylation; eye disease; glaucoma; tumour;  
KW inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcers;  
KW macular degeneration; inflammatory bowel disease; Crohn's disease;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antiarteriosclerotic; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200246454-A2.  
XX  
PD 13-JUN-2002.  
XX  
PF 06-DEC-2001; 2001WO-EPI4320.  
XX  
PR 06-DEC-2000; 2000DE-1061338.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Schacht O;  
XX  
DR WPI; 2002-500450/53.  
XX  
PT New nucleic acid fragments from chemically treated  
PT angiogenesis-associated genes, useful for determining methylation  
PT status, e.g. in diagnosis or treatment of cancer  
XX  
PS Claim 1; SEQ ID NO 105; 41pp + Sequence Listing; German.  
XX  
CC The invention relates to a nucleic acid (I) comprising a segment of 18  
CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)  
CC having sequences (ABQ66971-ABQ67178) or their complements. (I), also  
CC related oligomers, are used to evaluate the methylation status and/or  
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for  
CC diagnosis and treatment of eye diseases, proliferative retinopathy,  
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,  
CC diabetic retinopathy, macular degeneration caused by neovascularisation,  
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and  
CC Crohn's disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 7857 BP; 2110 A; 218 C; 2100 G; 3429 T; 0 other;

Query Match 3.4%; Score 36.8; DB 24; Length 7857;  
Best Local Similarity 54.4%; Pred. No. 16;  
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 829 TTGATCTTTGATATGACAGAGATGACAGAGTAACTTTTGAAAATTTCTAAAT 888  
Db 3507 TTAGATGTATTTTAAGACGGGAGATATTTTATTTTAGTATTTATTTTAAAAA 3566

QY 889 AGAATGCTTTGTCGAGCTAAAGCTTCTCTCTCTTTATTTTCTTTTGTAGATGAT 948  
Db 3567 AAAAAAATTTGTTAGTGTAGTGGTTATGTTTGTATTTTAGTATTTTGTGAAGAT 3626  
QY 949 TCGTACTAGGATAT 964  
Db 3627 AAGTAGTGGATTAT 3642

RESULT 47  
ABA03041/C  
ID ABA03041 standard; DNA: 2944528 BP.  
XX  
AC ABA03041;  
XX  
DT 05-FEB-2002 (first entry)  
XX  
DE Listeria monocytogenes EGD-e genome sequence.  
XX  
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
KW vitamin B12; bacterial infection; disease; ds.  
XX  
OS Listeria monocytogenes.  
XX  
PN WO200177335-A2.  
XX  
PD 18-OCT-2001.  
XX  
PF 11-APR-2001; 2001WO-FR01118.  
XX  
PR 11-APR-2000; 2000FR-0004629.  
XX  
PA (INSP ) INST PASTEUR.  
XX  
PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;  
PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;  
PI Daniels J, Goebel W, Krefit J, Kuhn M, Ng E, Vazquez-Boland JA;  
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
PI Maduenio E, De pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
PI Rose M, Voss H;  
XX  
DR WPI; 2002-010914/01.  
XX  
PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
PT and prevention of Listeria and related bacterial infections, and  
PT related polypeptides  
XX  
PS Claim 1; SEQ ID No 1; 192pp; French.  
XX  
CC The present sequence is the genome sequence of Listeria monocytogenes  
CC EGD-e. This sequence and fragments of this sequence are useful for  
CC selecting probes and primers for detecting genes in L. monocytogenes and  
CC related organisms, and to study genetic polymorphisms and other genomes.  
CC Proteins (AB847297-AB850149) expressed from the present sequence are  
CC useful for raising specific antibodies, identification of L.  
CC monocytogenes and related organisms, and for biosynthesis and  
CC biodegradation, especially biosynthesis of Vitamin B12. This sequence and  
CC proteins encoded by it are also useful for selecting compounds that  
CC regulate gene expression and cell replication and modulate L.  
CC monocytogenes-related diseases. In addition, this sequence and proteins  
CC encoded by it are useful in pharmaceutical and vaccines compositions for  
CC the treatment or prevention of infections by L. monocytogenes and related  
CC organisms.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;

Query Match 3.4%; Score 36.8; DB 24; Length 2944528;

Claim 1: SEQ ID 4186; 2337pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and

RESULT 49	
AAV22649/c	
ID	AAV22649 standard; cDNA; 2098 BP.
XX	
AC	AAV22649;
XX	
DT	17-AUG-1998 (first entry)
XX	
DE	Homo sapiens soluble kuzbanian (kuz) gene.
XX	
KW	kuzbanian; kuz; neurogenic; KUZ protein; neuronal partitioning;
KW	development; NOTCH protein processing; screening; cell function;
KW	signal transduction pathways; screening; receptor binding;
KW	metalloprotease; soluble; ds.
XX	
OS	Homo sapiens.
Key	Location/Qualifiers
FT	2..2098
CDS	/*tag= a
FT	/product= KUZ protein
FT	
PN	WO9808933-A1.
XX	
PD	05-MAR-1998.
XX	
PF	27-AUG-1997; 97WO-US15099.
XX	
PR	23-JUL-1997; 97US-0019390.
PR	29-AUG-1996; 96US-0019390.
XX	
PA	(REGC ) UNIV CALIFORNIA.
PA	{UYIA } UNIV YALE.
XX	
PI	Pan D, Rooke J, Dubin GM, Xu T, Yavari R;
XX	
DR	WPI; 1998-179428/16.
DR	P-PSDB; AAW56133.
XX	
PT	New KUZ polypeptides, members of the ADAM family of metalloprotease
PT	- useful in neural partitioning and development
XX	
PS	Claim 11; Pages 37-38; 58pp; English.
XX	
CC	The sequence is that of the human soluble (lacking a
CC	transmembrane domain) kuzbanian gene, a neurogenic gene.
CC	The KUZ protein that it encodes is involved in neuronal
CC	partitioning and development. It is also involved in processing
CC	CC

CC of the NOTCH protein by cleaving the C-terminal portion. The  
CC KUZ protein can be used to screen for compounds that alter  
CC binding of KUZ to its receptor or the cleavage of the NOTCH  
CC protein, hence acting to regulate NOTCH signal transduction  
CC pathways and regulate cell functions.

XX Sequence 2098 BP; 655 A; 353 C; 380 G; 710 T; 0 other;  
SQ Query Match 3.4%; Score 36.6; DB 19; Length 2098;  
Best Local Similarity 65.1%; Pred. No. 11;  
Matches 54; Conservative 0; Mismatches 29; Indels 0; Gaps 0;  
QY 816 AGCAACCTTCCTTTGATTTGATGATGACAGAGATGACAAAGAGTAAAGTTTGA 875  
Db 1136 ACCCAACAAGCTTTGAGGTTTGATTTTCAGAGAAGATGACCAAGAGTTCTTTTATTT 1077  
QY 876 AAAATTCTAAATAGAAATGCAT 898  
Db 1076 AGGCTACTGAAATATTTATTTAT 1054

RESULT 50  
AAH17997/c  
ID AAH17997 standard; cDNA; 2333 BP.  
AC AAH17997;  
XX 26-JUN-2001 (first entry)  
XX Human cDNA sequence SEQ ID NO:17797.  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX Homo sapiens.  
XX EP1074617-A2.  
XX 07-FEB-2001.  
XX 28-JUL-2000; 2000EP-0116126.  
XX 29-JUL-1999; 99JP-0248036.  
XX 27-AUG-1999; 99JP-0300953.  
XX 11-JAN-2000; 2000JP-0118776.  
XX 02-MAY-2000; 2000JP-0183767.  
XX 09-JUN-2000; 2000JP-0241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602  
XX full-length cDNAs defined in the specification, and for the detection  
XX and/or diagnosis of the abnormality of the proteins encoded by the  
XX full-length cDNAs -  
XX Claim 8; SEQ ID 17797; 2537pp + CD ROM; English.  
XX The present invention describes primer sets for synthesizing 5602  
XX full-length cDNAs defined in the specification. Where a primer set  
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
XX to the complementary strand of a polynucleotide which comprises one of  
XX the 5602 nucleotide sequences defined in the specification, where the  
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
XX of an oligonucleotide comprising a sequence complementary to the  
XX complementary strand of a polynucleotide which comprises a 5'-end  
XX sequence and an oligonucleotide comprising a sequence complementary to a  
XX polynucleotide which comprises a 3'-end sequence, where the  
XX oligonucleotide comprises at least 15 nucleotides and the combination of  
XX the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX Sequence 2333 BP; 699 A; 427 C; 416 G; 791 T; 0 other;  
SQ Query Match 3.4%; Score 36.6; DB 22; Length 2333;  
Best Local Similarity 65.1%; Pred. No. 11;  
Matches 54; Conservative 0; Mismatches 29; Indels 0; Gaps 0;  
QY 816 AGCAACCTTCCTTTGATTTGATGATGACAGAGATGACAAAGAGTAAAGTTTGA 875  
Db 567 ACCCAACAAGCTTTGAGGTTTGATTTTCAGAGAAGATGACCAAGAGTTCTTTTATTT 508  
QY 876 AAAATTCTAAATAGAAATGCAT 898  
Db 507 AGGCTACTGAAATATTTATTTAT 485  
Search completed: January 9, 2003, 10:51:58  
Job time : 2027 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 05:36:27 ; Search time 2610 Seconds  
(without alignments)  
12109.443 Million cell updates/sec

Title: US-09-438-185a-1\_COPY\_1199590\_1200675

Perfect score: 1086  
Sequence: 1 ttggcaagtacctaacaac.....gtctctgcagtagtgac 1086

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 150 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_m.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_on.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vl.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rdi.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_higo\_hum.\*  
40: em\_higo\_mus.\*  
41: em\_higo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
C 1	1086	100.0	1089	6	AX349679 Sequence
C 2	1086	100.0	10847	1	AE001685 Chlamydia
C 3	1086	100.0	22589	1	AE002240 Chlamydia
C 4	1086	100.0	325865	1	AF002548 Chlamydia
C 5	72.4	6.7	7218	6	166494 Sequence 14
C 6	46.8	4.3	151120	2	AC015814 Homo sapi
C 7	46.8	4.3	193804	9	AC084882 Homo sapi
C 8	46.6	4.3	1642	3	AF135186 Caenorhab
C 9	45.6	4.2	1531	3	AF119388 Caenorhab
C 10	45.6	4.2	41246	3	CEK08F8
C 11	45.6	4.2	109206	2	AC121238
C 12	45	4.1	1141	6	AX083744 Sequence
C 13	45	4.1	172246	9	AC024941 Homo sapi
C 14	45	4.1	180563	2	AC034170 Homo sapi
C 15	44.8	4.1	12029	3	AE001431 Plasmid
C 16	44.2	4.1	783	9	HSA343706
C 17	44.2	4.1	234112	3	PMAL4P2
C 18	44	4.1	167257	9	AC008008 Homo sapi
C 19	44	4.1	198230	9	AC092662 Homo sapi
C 20	43.8	4.0	1141	6	AX083744 Sequence
C 21	43.4	4.0	6089	6	AX344846
C 22	43.2	4.0	88966	2	AP001943
C 23	43.2	4.0	146671	9	AC093810 Homo sapi
C 24	42.8	3.9	158066	9	AC107307 Homo sapi
C 25	42.8	3.9	159716	2	AC025807 Homo sapi
C 26	42.8	3.9	186158	9	CNS018F0
C 27	42.8	3.9	188385	2	AC112384 Rattus no
C 28	42.4	3.9	159035	9	AC092584
C 29	42.2	3.9	2706	1	AF268060
C 30	42.2	3.9	126010	9	AC010342
C 31	42.2	3.9	141635	2	AC108007
C 32	42.2	3.9	132168	2	AL807375 Mus muscu
C 33	42.2	3.9	202801	2	AL831718 Mus muscu
C 34	42.2	3.9	266050	1	AP000988 Sulfolobu
C 35	42	3.9	98697	9	AC004854 Homo sapi
C 36	42	3.9	124767	2	AC108637
C 37	41.8	3.8	363	1	MC445
C 38	41.8	3.8	76141	2	AC024393
C 39	41.8	3.8	115666	2	AC105744 Homo sapi
C 40	41.8	3.8	126371	9	AL391883 Human DNA
C 41	41.8	3.8	156060	2	AC004153 Plasmid
C 42	41.8	3.8	157515	2	AL391598 Homo sapi
C 43	41.8	3.8	194247	2	AL358253 Homo sapi
C 44	41.8	3.8	200279	2	AC109165 Mus muscu
C 45	41.6	3.8	27785	2	AC116978 Dictyoste
C 46	41.6	3.8	69674	9	AL136130 Human DNA
C 47	41.6	3.8	103337	2	AC115405 Rattus no
C 48	41.6	3.8	110000	2	Continuation (2 of
C 49	41.6	3.8	148549	2	AL353694_1
C 50	41.6	3.8	148549	2	AL354818 Homo sapi
C 51	41.4	3.8	141885	2	AC046180 Homo sapi
C 52	41.4	3.8	181200	2	AC026518 Homo sapi
C 53	41.4	3.8	345777	2	AC099365 Rattus no
C 54	41.2	3.8	14514	2	AC078629 Homo sapi
C 55	41.2	3.8	147192	2	AC115610 Dictyoste
C 56	41.2	3.8	153133	2	AC116925 Dictyoste
C 57	41.2	3.8	170668	2	AC106207 Rattus no
C 58	41.2	3.8	197127	2	AC091787 Oryz sat
C 59	41.2	3.8	349980	6	AC098253 Rattus no
C 60	41	3.8	1202	8	AF308778 Arabidops
C 61	41	3.8	9504	6	AX323794 Sequence
C 62	41	3.8	40392	3	U21308 Caenorhab
C 63	41	3.8	137830	2	AC024346 Homo sapi
C 64	41	3.8	151225	9	AC012625 Homo sapi
C 65	41	3.8	152679	2	AC016082 Homo sapi

• 五



```
/db_xref="GI:4377383"
/translation="MSKIVKFGGTSLATAENICLVCDIICKDKRPSFVVVSAIAGVTD
LIVDCSSLREREVEKTEGHEIIVKNLAIPFPVSTWTSRLPYLQHLIETDLPF
ARILSGEDIIASIVRAVCTRWGDLGFLEARSVILDDSVRASPNLDMKAHWHOL
EINQPSYIOGFIQSGLGTGVLGGGSDYSATILAEALRATERYIVDVGNYTMD
KVIISDAQRIPELFSEMQNLASFGAKVLYPPMLPCMRAGIPFVTSFDPKGGTW
VAVDSYSEYPRIKALSLSQYQFCSDVYTVLGGGLEILIEHSGIDPELMIAQ
NUNVGFMDDDIISOEAQELVDVLSLSSVTRLHHSVALITMIGNLSSPKVVSITIE
KRGVQGVDFVFCQSSMALSFVVASLAEGLIEELNHDVVKQKAIATV"
5162..6037
/gene="dapa"
/rota="Cpnl050"
5162..6037
/gene="dapa"
/codon_start=1
/transl_table=1
/product="dihydrodipicolinate Synthase"
/protein_id="AAD19187.1"
/db_xref="GI:4377384"
/translation="MHLLTATVPFPNGTIDFASLERLLSFQDAGVGVLLGSGTE
GLSLTKKQALICFACDLQKVPFLFVGTSGTLLEEDVDMWHFCNDLPISGFLMTPI
YTKPLKQGLIWFCAVLNAAKHPAILYNIPSRATPLVLDTVKALAHHPQLGFKDSG
GSVEFQSYKFIAPHIQIGCDGVFWSEMAACGAHLISVLSNWAPPEARAYVLNPOE
QDYSLMMETCRWYVTTNPIGIKAILAYKKAITHAQLRLPLSIEDFLENVSPVES
MLAWPKLRYSVESYS"
6198..6509
/gene="Cpnl051"
6198..6509
/gene="Cpnl051"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAD19188.1"
/db_xref="GI:4377385"
/translation="MTPKSIQQLHLTKIDIPVRKISPVTTKKSSFROSLLRLFLELFW
MFLYICIRSIHFCHVIAIFICRGLILFLATLFLSMICILHFTITLPWICKEDPRIIRN
K"
6653..7408
/gene="Cpnl052"
6653..7408
/gene="Cpnl052"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAD19189.1"
/db_xref="GI:4377386"
/translation="MKNYSREKIKSALRICSSVCITVFRNFSLSYDKIFYLSVCYV
FNGPNSIGCRSFCFPRGKTEVETKEVKKDEIRPSLEGNDPVKVAESPFRKAALE
SLSSQSSIGNLCALSNFLDSQMLSRNFSKEIMGSIIFTRSKTCDAGESEPFRTAGG
VLAGRPLKAGLSYELGTAGLQGLKQVSDSHRTATSSILSVHGSVWTRPLSCTKY
IVGKARPLLFERLTSVDRDLKKRFLRECD"
7389..7940
/gene="Cpnl053"
7389..7940
/gene="Cpnl053"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAD19190.1"
/db_xref="GI:4377387"
/translation="MNFAKIDHNHLYLTCIGDIGVACPILSTDCPLNPYSEKASHEVLV
YSKFCRISGEPSRLATSGNTDYYISVLPIGRYEVTFSPGRHDFNDIMHVAPKIGAV
LHGKREAKEIPGTSKQDYAFPSLTARESMLSEKLAMTPOVSEVIONCYSOCTKVTKT
NLKEQYRHLSHNTGFELSVKSAF"
8270..10705
/gene="Cpnl054"
8270..10705
/gene="Cpnl054"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAD19191.1"
/db_xref="GI:4377388"
```

```
/translation="MQVLLSPQLPPPHQHSVSGSISSPKRLVLAIATFLVFGMLLIISG
ALFLTIGLGAATISFGIGLGSALGVLMSIGLCLLVKREIPYRPEEIPEGVSL
APSEEPALQAAITLAQPKELDQDIDIEVACFKLKLKDKYKRSFSLNDKAKREL
VFDFVVEDTLISFIFELQIVAQEGWDLNFIINGRSLMNTAASESLDPLHPSKRLGYL
PSDVRGEGLSKSAKEIVARLMSLHCEIHKVAFDRNSYMAEKAFKALGALAEVS
YSLTQSYRDKFLSESAKIPWNGHITWLRDDAKSGCAKKKLRDAERWKKFKKAVFW
VEGSGDFDNLGLDGTVLDYRQERDEITTHELYETTKTLRLHRSKALAKTTFE
KRSKNKCAVEANRRLKYVRWDYQEFQKAGERLEKLHLYLPEVSYIRENKIOE
TRSNLEKAYATEPNYRCCVREQEDYMKEREKEAREFERGRKILSPESLESSQPD
HGLNFESEKLEGLEHILKLOKATVAENKILSDAESRLIYFEDVEMPORIEIE
KTLRMAELPLPTKAFKAGSOVNSCAEMLEKWKPYCKESLAYVTSKRLVSLDEDL
RRAYTECOKRFQDGLSESEVRAGREQLRERIOEFETQGLDLYVEKELLVCSRLANTE
CDCYSGYKKEAPPGRKFTAQHIDELIYRVQSRWMTSRLREGVQACNKKMAGLSE
EDKLVKEEYIWLIRERKNKRLVQIKVATQORVAAAFESIEVPEIPAPKEKPSLL
DKARSFTREDHT"
BASE COUNT 3261 a 1883 c 2436 g 3267 t
ORIGIN
Query Match 100.0%; Score 1086; DB 1; Length 10847;
Best Local Similarity 100.0%; Pred. No. 4e-244;
Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGGCAAAAGTACCTCAAAACGAGAAAGATATTCTCTTGATTGTAAGGGGAATTGATT 60
Db 826 TTGGCAAAAGTACCTCAAAACGAGAAAGATATTCTCTTGATTGTAAGGGGAATTGATT 885
QY 61 TAACAGACCTTGGTCGAGCATCCATCTTAATTTGAAGTGAGTCTACCACTTCATCAAA 120
Db 886 TAACAGACCTTGGTCGAGCATCCATCTTAATTTGAAGTGAGTCTACCACTTCATCAAA 945
QY 121 ATGCTTATTAAGAAATAAGTCTCTTGTGGAGTGATGATTTGAAGGGAAGCAATAAT 180
Db 946 ATGCTTATTAAGAAATAAGTCTCTTGTGGAGTGATGATTTGAAGGGAAGCAATAAT 1005
QY 181 CTGATCCAAATTTAAAGGGAGAACACGTACGTTATCAATGAAGAGCGTGTCACAGTTCCGT 240
Db 1006 CTGATCCAAATTTAAAGGGAGAACACGTACGTTATCAATGAAGAGCGTGTCACAGTTCCGT 1065
QY 241 AGGAGAACTGATAAGAACGGCTCCATATGCTTTTCTCTTCATGCTTTTCAATAAGTCC 300
Db 1066 AGGAGAACTGATAAGAACGGCTCCATATGCTTTTCTCTTCATGCTTTTCAATAAGTCC 1125
QY 301 GCTTTCACACAGTAAAGAACGGCTCCATATGCTTTTCTCTTCATGCTTTTCAATAAGTCC 360
Db 1126 GCTTTCACACAGTAAAGAACGGCTCCATATGCTTTTCTCTTCATGCTTTTCAATAAGTCC 1185
QY 361 TATGGTGTGTTTTTACTAGSAGAGCTTGACTTTTCTATGACTTTAGTGAAGAGTCT 420
Db 1186 TATGGTGTGTTTTTACTAGSAGAGCTTGACTTTTCTATGACTTTAGTGAAGAGTCT 1245
QY 421 TCCCATGTTTATGAATAATTCAGAAATGAGGATGTAGAAGCCAGGCGAGTCCCTAA 480
Db 1246 TCCCATGTTTATGAATAATTCAGAAATGAGGATGTAGAAGCCAGGCGAGTCCCTAA 1305
QY 481 AAGTTCATGATGATCAGGAGTTAGGAGAAATATCTTTATGAGAGTTCCGATTAC 540
Db 1306 AAGTTCATGATGATCAGGAGTTAGGAGAAATATCTTTATGAGAGTTCCGATTAC 1365
QY 541 AGAGCAATGGGAAGTAACGGTCTCGCAACAGACAGCAAGTATTTGATGGGGAGGTACAAA 600
Db 1366 AGAGCAATGGGAAGTAACGGTCTCGCAACAGACAGCAAGTATTTGATGGGGAGGTACAAA 1425
QY 601 TCCCTGAAACGGGATAATAGGAAAAATGTGTTCTTAATTCGAAGAATTTAATGACTGCTTG 660
Db 1426 TCCCTGAAACGGGATAATAGGAAAAATGTGTTCTTAATTCGAAGAATTTAATGACTGCTTG 1485
QY 661 ATGGTCTAAGAAATCAGAAAGGAGACCGGAAAGCTTCTAGTAATCAAGAAAAAATCTTG 720
Db 1486 ATGGTCTAAGAAATCAGAAAGGAGACCGGAAAGCTTCTAGTAATCAAGAAAAAATCTTG 1545
QY 721 ACAATAGCTTTTCCATAGACAAACCGTGAGGAAGGAGTGCATACCAATTAATTCGATT 780
Db 1546 ACAATAGCTTTTCCATAGACAAACCGTGAGGAAGGAGTGCATACCAATTAATTCGATT 1605
```

QY 761 AAAAAAGGAGTCTTACAAAGCCATATCAGTGGAGCAACCTTCCCTTGATCTTITGA 840  
|||||  
Db 1606 AAAAAAGGAGTCTTACAAAGCCATATCAGTGGAGCAACCTTCCCTTGATCTTITGA 1665  
|||||  
QY 841 TAATGACAGAAGATGAACAAAGAGTAAGTTTTTGAAGAAATTCATAAATGAAATGCAFTT 900  
|||||  
Db 1666 TAATGACAGAAGATGAACAAAGAGTAAGTTTTTGAAGAAATTCATAAATGAAATGCAFTT 1725  
|||||  
QY 901 GTGTCGAGCTAAAGCTTGCTTCTCTTTTATTTCTTTTGTAGAAATGATTCGGTAGTAGGA 960  
|||||  
Db 1726 GTGTCGAGCTAAAGCTTGCTTCTCTTTTATTTCTTTTGTAGAAATGATTCGGTAGTAGGA 1785  
|||||  
QY 961 ATATGGGCTCGAGTATGACAGCTTGAGGTGATTCGCTGCTTCTGGAAGAACAGGAAAGTGA 1020  
|||||  
Db 1786 ATATGGGCTCGAGTATGACAGCTTGAGGTGATTCGCTGCTTCTGGAAGAACAGGAAAGTGA 1845  
|||||  
QY 1021 TTGCTCAGCTTTAGAGCAATCTTCAGAAATATCTTTGGGTCCAGGTTCTCTCGCAGTA 1080  
|||||  
Db 1846 TTGCTCAGCTTTAGAGCAATCTTCAGAAATATCTTTGGGTCCAGGTTCTCTCGCAGTA 1905  
|||||  
QY 1081 GTGCAC 1086  
|||||  
Db 1906 GTGCAC 1911  
|||||

RESULT 3  
AE002240/c  
LOCUS  
DEFINITION Chlamydomophila pneumoniae AR39, section 68 of 94 of the complete genome.  
ACCESSION AE002240 AE002241 AE002161  
VERSION AE002240.2 GI:8163502  
KEYWORDS  
SOURCE Chlamydomophila pneumoniae AR39.  
ORGANISM Chlamydomophila pneumoniae AR39  
REFERENCE  
1 (bases 1 to 22589)  
Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Winn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.  
Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39  
Nucleic Acids Res. 28 (6), 1397-1406 (2000)  
JOURNAL  
MEDLINE  
PUBMED 10684935  
REFERENCE  
1 (bases 1 to 22589)  
Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Winn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.  
Direct Submission  
Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
COMMENT  
On or before Jun 1, 2000 this sequence version replaced gi:7189730, gi:7189720.

FEATURES  
source  
1. .22589  
Location/Qualifiers  
/organism="Chlamydomophila pneumoniae AR39"  
/strain="AR39"  
/db\_xref="taxon:115711"  
/note="synonym: Chlamydia pneumoniae AR39"  
complement(161..955)  
/gene="Cp0805"  
complement(161..955)  
/gene="Cp0805"  
/note="similar to GB:D10483 SP:P04036 GB:M10611 PID:145710  
PID:216458; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11

gene  
CDS  
/product="dihydrodipicolinate reductase"  
/protein\_id="AAF38603.1"  
/db\_xref="GI:7189721"  
/translation="MGSSMHVGVIGSGRTGKIVSALEQSESYTLPGFSRSSALT  
LVYTAHNDVLVDFSHPLTKEVVAHLILSPKPLIIGTTGPGCKEADHSLRELT  
FOVYCPNAGLXGAYTHKELWMLISQLCNQPDRIRETHRRYKDKSLSGTQDL  
LDITIQ QVKDEGGEREYEGQDSSKKTILEVQSSRVDIPGEHEVAFIS  
SGEOLLVRHTVFSRN VFGKILSLDLWLKTLNPQGLYSGDTLELVLRNEHCLLKKT  
DTH" 832..1920  
/gene="Cp0806"  
832..1920  
/gene="Cp0806"  
/note="bioplerin-dependent aromatic amino acid hydroxylase  
family protein; identified by match to PFAM protein family  
HMM PF00351"  
/codon\_start=1  
/transl\_table=11  
/product="bioplerin-dependent aromatic amino acid  
hydroxylase family protein"  
/protein\_id="AAF73705.1"  
/db\_xref="GI:8163503"  
/translation="MHYCERLDPKYILKIALKRLQSLFFQNSQSLQRAYSTPYSY  
YRIILQENKEKQALARKHCISILEFFKNLLEVLHLSLKNQREGCTDMVAVSTP  
FNLMWRLSSRFSLKSYCPFFLDYLEAGLLSDFLDHQAIVKFFELHFSYTPV  
SGFVAPHOYLSLQDQRYFPASVMTLDKDNFSLTDLHDLGLHVPWLLHPSSEFF  
INMGRLFTKIEKVOALPSSKQRTQTLQSNLIAI VRCEFWTVESGLIENHEGRKAYGA  
VLISSPQELGHAFIDNVRLPLLELDQIIRLPENTSTPOETLFSIRHDFELVELTSKLE  
WMLDGLLESIPLYNQEKYLSGFEVLQ" 1911..2609  
/gene="Cp0807"  
1911..2609  
/gene="Cp0807"  
/note="conserved hypothetical protein; identified by  
Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF38604.1"  
/db\_xref="GI:7189722"  
/translation="MPMNTSHRKTLPVSLSTFLLVLSNLVSSKLPTTFNFPI  
IFGHCILPLFLSDVNVNEIFGKARVMSAFIANLLASSIVQIFMFFPVASPEM  
QWAMHCLDPLRPLASLAFIVSQDLIDVLYTFKNTNSSLKNSGTSWISQI  
PTTFIVDCILYFGMGLSFPOTLIMFYIYKIFCVLTTPFLYLAVENTIRKELGMP  
STKIANTVPLINQP" 2816..3811  
/gene="Cp0808"  
2816..3811  
/gene="Cp0808"  
/note="similar to GB:J04423 SP:PI2996 PID:145425  
PID:490219 GB:U00096; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="biotin synthase"  
/protein\_id="AAF38605.1"  
/db\_xref="GI:7189723"  
/translation="MREFTVMSLSDIREIYHTPVFLIHKANILRSNLSLQTC  
YLIISIKTGCVEDCAYCAOSSRYHTVTPPMKIVDVVERAKRAVELGATRVCLGAA  
WRNAKDDRYEDVLAMVKSITDLCAEVCCAGMLSEQAQKLYDAGLYAVNHNLDSSP  
EFTVTLITRSYEDRLTLDVYKSGISCTCCGICVCMGESDEEDIKLHLVLAIRHIP  
EFTVNLWLPIDGCTPLQDQPDISEFVRVLIATARVVPFRSNVRLAAGRAFLTVQQT  
LCFLAGANSIFYGDKLLIVENDIDEDAEIMIKLLGLIPRPSGIERGNPCYANNS"  
3790..4938  
/gene="Cp0809"  
3790..4938  
/gene="Cp0809"  
/note="similar to GB:AL009126; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="8-amino-7-oxononanoate synthase"  
/protein\_id="AAF38606.1"  
/db\_xref="GI:7189724"

gene  
CDS

/translation="MLCQFLEIALARRKSKHTYRSLSINSHLIDFTNDYLGFASSP  
ELRKYIKLHAIESLGLVATGSRLLTGHLSOLCORIEQLAAHYHNFESCLIFWTGYTANL  
GLYALATDQDRIHLIDLY IHASIVDGLIRLSKQSPFNHNDLHLEKRLASHLGRTE  
VCVESVYSGVAPLOAA ISELCEYSAYLIVDEAHAVGVFGDQEGELVSGALQDQV  
LTVTYFKALGTGACAAIGSSILKDYLLNFCRPFYITTAOPPHALTAIELAYEHNOR  
AFNOREHLSALIHFRKAKQNLGLOLMKONTTPTOSICVSGSHPARQAAIQNSGY  
DVRPIVSTVRORELLRICLGHAFNTKNEIDLHLHTLEQIFLCNVSSL"  
4917..5576  
/gene="CP0810"  
CDS  
4917..5576  
/gene="CP0810"  
/note="similar to SP:P53558 PID:1277028 PID:2293186  
GB:AL009126; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="dethiobiotin synthase Biob, putative"  
/protein\_id="AAF38607.1"  
/db\_xref="GI:7189725"  
/translation="MQRITIVGDTGVGKTIYSAILARALNAEYWKPIQAGNLSNDS  
NIVHLSGAYCHPEAYRLHRLPLSPHAKAQIDNYSIEESHICAPKTTSLNLIETSGGFL  
SPCTKRLQGVDFSSWSCWILVSOAYLGSINHTCLTVEAMSRNLNLIGMVVNGYPE  
DEEHLTQEIKLPIIGTILAKEKEITKTIISCYAEQWKEVWTSNHOGIQGVSGTPSLN  
H"  
5513..6784  
/gene="CP0811"  
CDS  
5513..6784  
/gene="CP0811"  
/note="similar to GB:J04423 SP:PI2995 PID:457106  
PID:490217 GB:U00096; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="adenosylmethionine-8-amino-7-oxononanoate  
aminotransferase"  
/protein\_id="AAF38608.1"  
/db\_xref="GI:7189726"  
/translation="MDKOSGSGCINHPHTQSALDSTPIKIVREGAYLYAESGTRY  
LDAISWNCNHLGHGHPYIKLCEQAQKLEHVFIFANFHEPALEKSLAPLLEGL  
ERFFSDSGNSIEMAKIVAYQYQNKAKSHFVGLSNAYHGDFTGAMSTAGTSPT  
VFPHDLFSTIAAPYIGKEELAIAQATFVSESNIAFIYEPFLQAGAGMLWYNE  
GLKEILKRLKAGVLCIADBELTGFGRGTPLFASETDIPDDIICLSKGLTGYPLA  
LVTWTEHDAFVSQDRMKALLHGTFTGNPLGCSAAALSDLTSLSPCLQOQOMIBR  
CHQFEQAEHSLMORCEVLGTLALDYPAEATGYFSQVDRHLNRFLERGVLLRPLGN  
TLYVLPPYCIQEDRLIYSHLODALCLOP"  
6766..7437  
/gene="CP0812"  
CDS  
6766..7437  
/gene="CP0812"  
/note="hypothetical protein; identified by Glimmer2;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF38609.1"  
/db\_xref="GI:7189727"  
/translation="MSTITVVKHFIHTASRWEPLKIEVASNVHQAQWINTLSPLNENG  
AKKISASHPTEVKEEVKHAABEFRRGHYLTQISRISETSLPDYTSKNLLGLLTK  
YTLHLDLTRCVLENSYSLSGOTLTAAYILVYAIELRASELYPLHYDLLEKQSK  
ITVKSIILEEOGHLOEMERELKDPHPGELLGYACQFEGELCLOFVERLEQMIDPPSS  
TTFK"  
7508..8845  
/gene="CP0813"  
CDS  
7508..8845  
/gene="CP0813"  
/note="3-phosphoshikimate 1-carboxyvinyltransferase;  
identified by match to PFAM protein family HMM PF00275"  
/codon\_start=1  
/transl\_table=11  
/product="3-phosphoshikimate 1-carboxyvinyltransferase"  
/protein\_id="AAF37306.1"  
/db\_xref="GI:8163504"  
/translation="MLYKYSPSSVYGNAPIPSSKSHLTRALLHVASVECKSIYNYL  
DSPDPEAMICKAKMGASIKKFPQILEIVGNFLAIFPKYTLIDAGNSGIVLRFMTALA

gene  
CDS

CVFSKEITVTGSSQLQRRPMAPLQALNRFASPHFSSDKSVLPFTMSGPLRSYSDV  
EGSDQFASALAVACSLAEGPCSFTHIEPKRPMFDSLWMLKHLPLPYSCSDTYSP  
PGSSHPQGSFYHVTGDFSSAAFTAAALLSKSLQPIRLNLDILDQDKIFSLMON  
LGASIOYDNEELVLPSPSSGSDMDGCDIDALPILVLCFADSPSHLYNARSADK  
ESDRILATTEELQKMGACIQTPHGLNPSPLYGVLSDHDDHRLTAMALTAAALYAS  
GDSRLHNTACVRKTPPNFVOTLINMEARIECHDNYSWNSHTRKRYFARESFG  
8781..9311  
/gene="CP0814"  
CDS  
8781..9311

Query Match 100.0%; Score 1086; DB 1; Length 22589;  
Best Local Similarity 100.0%; Pred. No. 3.7e-244;  
Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGGCAAAAGTACCTCAAAACCCAGAAAGATATTTCTTGATTGTAAGGGGAATTGATTC 60  
DB 1917 TTGGCAAAAGTACCTCAAAACCCAGAAAGATATTTCTTGATTGTAAGGGGAATTGATTC 1858  
QY 61 TAACAGACCTTTGGTCGAGCATCCATTTCTAATTTTGAAGTGAGTTCTTACCAGTTTCATCAA 120  
DB 1857 TAACAGACCTTTGGTCGAGCATCCATTTCTAATTTTGAAGTGAGTTCTTACCAGTTTCATCAA 1798  
QY 121 ATGTCTTATTGAAATAAAGTCTCTTGTGGAGTTGATGTATTGAAGGAAGACGAATAAT 180  
DB 1797 ATGTCTTATTGAAATAAAGTCTCTTGTGGAGTTGATGTATTGAAGGAAGACGAATAAT 1738  
QY 181 CTGATCCAAATTTAAAGGGAGACACAGTACGTTTATCAATGAAGCGTGTCCAGTCTCG 240  
DB 1737 CTGATCCAAATTTAAAGGGAGACACAGTACGTTTATCAATGAAGCGTGTCCAGTCTCG 1678  
QY 241 AGGAACTGATAGAACGCGTCCATATGCTTTTCTTCTTCATGGTTTTCATTAAGTCC 300  
DB 1677 AGGAACTGATAGAACGCGTCCATATGCTTTTCTTCTTCATGGTTTTCATTAAGTCC 1618  
QY 301 GCTTTCAACAGTAAACCAAGCAGCGTACAATAGCGATCAGATTGCTTTGTAGGGTTG 360  
DB 1617 GCTTTCAACAGTAAACCAAGCAGCGTACAATAGCGATCAGATTGCTTTGTAGGGTTG 1558  
QY 361 TATGGTTGTTTTTTTACTAGSAGAGAGCTGTACTTTTTCTATGACTTTTAGTGAAGAGTCT 420  
DB 1557 TATGGTTGTTTTTTTACTAGSAGAGAGCTGTACTTTTTCTATGACTTTTAGTGAAGAGTCT 1498  
QY 421 TCCCATGTTTATGAAAATTCAGAAAATGAGGATGATAGAAGCAAGGACGTCGCCATA 480  
DB 1497 TCCCATGTTTATGAAAATTCAGAAAATGAGGATGATAGAAGCAAGGACGTCGCCATA 1438  
QY 481 AAGGTCATGGATGAGATCAGGAGTTAAGGAGAAATTTATCTTTATCGAGAGTTCCGATTAC 540  
DB 1437 AAGGTCATGGATGAGATCAGGAGTTAAGGAGAAATTTATCTTTATCGAGAGTTCCGATTAC 1378  
QY 541 AGAGCAATGGAAAGTAACGGTCTGCAACAGACAGATATTGATGGGAGCTACAAA 600  
DB 1377 AGAGCAATGGAAAGTAACGGTCTGCAACAGACAGATATTGATGGGAGCTACAAA 1318  
QY 601 TCCCGAAACGGGATTAATAGGAAAATATGTTTCTTAATTCGAAAGATTTAATGACTGCTTG 660  
DB 1317 TCCCGAAACGGGATTAATAGGAAAATATGTTTCTTAATTCGAAAGATTTAATGACTGCTTG 1258  
QY 661 ATGTGCTTAAGAAATCAGAAAGGAGACCGGAAAGCTTCTAAGTAATCAAGAAAAATCTTGG 720  
DB 1257 ATGTGCTTAAGAAATCAGAAAGGAGACCGGAAAGCTTCTAAGTAATCAAGAAAAATCTTGG 1198  
QY 721 ACAATAGCTTTTCCATAGAGAAAACCGTGAAGAAAGAGTTCGATACCAATAAATTCGGATT 780  
DB 1197 ACAATAGCTTTTCCATAGAGAAAACCGTGAAGAAAGAGTTCGATACCAATAAATTCGGATT 1138  
QY 781 AAAAAGGAGCTGCTTACAACGCCATATCAGTGAGGACCAACCTCCCTTCGATCTTTGA 840  
DB 1137 AAAAAGGAGCTGCTTACAACGCCATATCAGTGAGGACCAACCTCCCTTCGATCTTTGA 1078  
QY 841 TAATCAGACAGATGACAAAGAGTAGTTTGTGAATATCTTAATATAGAAATGCAATTT 900  
DB 1077 TAATCAGACAGATGACAAAGAGTAGTTTGTGAATATCTTAATATAGAAATGCAATTT 1018

```
QY 901 GTGTCGAGCTAAAGCTTGCTTCTCTTTATTTCTTTGTTAGAAATGATTCGGTAGTAGGA 960
|||||
Db 1017 GTGTCGAGCTAAAGCTTGCTTCTCTTTATTTCTTTGTTAGAAATGATTCGGTAGTAGGA 958
|||||
QY 961 ATATGGGGTCCAGTATGCACGTTGGAGTGAATGGCTGTTCTGGAAGAACAGGAAAGTGA 1020
|||||
Db 957 ATATGGGGTCCAGTATGCACGTTGGAGTGAATGGCTGTTCTGGAAGAACAGGAAAGTGA 898
|||||
QY 1021 TTGTCCTAGCTTTTAGAGCAATCTTCAGAAATATACCTTTGGGTCCAGGGTCTCTCGCAGTA 1080
|||||
Db 897 TTGTCCTAGCTTTTAGAGCAATCTTCAGAAATATACCTTTGGGTCCAGGGTCTCTCGCAGTA 838
|||||
QY 1081 GTGCAC 1086
|||||
Db 837 GTGCAC 832

RESULT 4
AP002548 325865 bp DNA linear BCT 25-MAY-2002
Chlamydothila pneumoniae J138 genomic DNA, complete sequence,
section 4/4.
ACCESSION AP002548 AB033788 AB033789 AB033790 AB033791 AB033818 AB036083
AB036084 AB036085 AB036086 AB036087 AB036088 AB036089 AB038350
AB038351 AB038352 BA000008
AP002548.2 GI:10176694

Chlamydothila pneumoniae J138 (strain:J138) DNA.
Chlamydothila pneumoniae J138
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydothila.
REFERENCE 1
AUTHORS Shirai.M., Hirakawa.H., Ouchi.K., Tabuchi.M., Kishi.F., Kimoto.M.,
Takeuchi.A., Nishida.J., Shibata.K., Fujinaga.R., Yoneda.H.,
Matsushima.H., Tanaka.C.; Furukawa.S., Miura.K., Nakazawa.A.,
Ishii.K., Shiba.T., Hattori.M., Kuhara.S. and Nakazawa.T.
TITLE Comparison of outer membrane protein genes omp and pmp in the whole
genome sequences of Chlamydia pneumoniae isolates from Japan and
the United States
J. Infect. Dis. 181 Suppl 3, S524-S527 (2000)
20298986
REFERENCE 2
AUTHORS Shirai.M., Hirakawa.H., Kimoto.M., Tabuchi.M., Kishi.F., Ouchi.K.,
Shiba.T., Ishii.K., Hattori.M., Kuhara.S. and Nakazawa.T.
TITLE Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and CWL029 from USA
Nucleic Acids Res. 28 (12), 2311-2314 (2000)
20330349
REFERENCE 3 (bases 1 to 325865)
AUTHORS Shirai.M.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2000) Mutsunori Shirai, Yamaguchi University
School of Medicine, Department of Microbiology, 1-1-1
Minamigoshu, Ube, Yamaguchi 755-8505, Japan
(E-mail:mshirai@po.cc.yamaguchi-u.ac.jp, Tel:81-836-22-2227,
Fax:81-836-22-2415)
COMMENT On or before Sep 15, 2000 this sequence version replaced
gi:6172302, gi:6172304, gi:6172306, gi:6172308, gi:6172400,
gi:6635182, gi:6635184, gi:6635186, gi:6635188, gi:6635190,
gi:6635192, gi:6635194, gi:8547442, gi:8547447, gi:8547451,
gi:8979170.
FEATURES Location/Qualifiers
source 1..325865
organism="Chlamydothila pneumoniae J138"
strain="J138"
/db_xref="taxon:138677"
/note="synonym:Chlamydia pneumoniae (strain J138)"
21..1318
/gene="CPj0797"
221..1318
/gene="CPj0797"
/codon_start=1
/transl_table=11

product="hypothetical protein"
/protein_id="BAA99005.1"
/db_xref="GI:8979171"
/translation="MSKKIKVLGHLTLCTLFRGVLCAALSNIGYASTSOESPYQKSI
EDMGYFTDLELLSKDEGWESEAHATSGNGSRIVGASGAGQSVTAVIWESHILKHLGT
LGGASAEIGISKDDGEVVGWSDTREGYTHAFVFDGDMKDLGTLGATYSVARGVSGD
GSIIVGYSATARGEDYGQWGVKWKGIKOLKLLPOLSEANAISETGVIVGSGE
ISRNHIVAKNNKNNAVYSLGTGLGGSVAAEAISANGKVIWGSTNNGETHAFMKHDE
TMHDLITLGGFSVATGVSADGRAIVGFSVAKTGEIHAFYAYEGEMEDLTTLGGEAR
VFDLSSGENDLIGSIKIDAGAEAYLFHLHK"
1467..2480
/gene="CPj0798"
1467..2480
/gene="CPj0798"
/codon_start=1
/transl_table=11
product="hypothetical protein"
/protein_id="BAA99006.1"
/db_xref="GI:8979172"
/translation="MKKTCQNTSRISGVVFSVLELTOTLPAGHFIDIGTSLYSW
ARGVSGDGRVVGVEGYNFAYVDGEKFLLEGLYPERSEALYFKASYDGSVILGISDOD
PSCRKAVKNGALVDLGIIFSEGMSFARGVSSDGIIVGCLYSDTETNFAVRWDETG
MYVLNLPEDRHSACWDASEDSVIVDAMGSEETAKAVYWKDGEHLLSNIGAKRS
SAHAYSKDGSFIVGEFISEENEHAFYVYHNGVIKDIGTLGGDI SVATGYSRDSKVIIVG
HSTRYDGEYRAFVKYVDGRMIDLTLGGSASFAGVSDDGKTIIVGKFEITELGECHAFIY
LDD"
complement(2558..3607)
/gene="CPj0799"
complement(2558..3607)
/gene="CPj0799"
/codon_start=1
/transl_table=11
product="hypothetical protein"
/protein_id="BAA99007.1"
/db_xref="GI:8979173"
/translation="MAALIKQLRSMLSQSSLLMVLFSYLSGICYVITDKPEDDFHS
SSAKWDHNGKTLTSLNSNKASAKAVGTGATTGVIKDTWTSYAVVRNWTGKEL
PTSSWVKSKANGISDGSIIAGIVENELSQSFATWKNEMYLLPSTWAVQSGKAYGI
SPDGSVIVSKADMSRTFAVGTGHEAOVLPGVMAKVSANVSANGSIIVGSDQA
SCILYAVKNEGNTITHLCTLGYSATIAKAVSNNKVIIVGRSTYYGEVHAFCRNGVM
SDIGTLGGSYSAKGSVATGKVIIVGSTMTTANGKLHAFKYVGGRMIDLGLCEYSWKACAN
AVSIDGELIIVGQSE"
complement(3867..5153)
/gene="eno"
complement(3867..5153)
/gene="eno"
/codon_start=1
/transl_table=11
product="enolase"
/protein_id="BAA99008.1"
/db_xref="GI:8979174"
/translation="MFEAVIADIQARELLDSRGVPTLHVKVTTSTGCVCEARVPSCAS
TGKKEALFRTDPSRYQGVKQVLAQVKNVREILFPLVKGCVSYEOSLIDSLMDSGDS
PNKTELGANAILGVSLATAHAAALRRPLRYLGGCFACSLPCPMNLINGMHADN
GLGQEFMIRPIGASSIKKAVNMGADVPHTLKLLHBERGLSTGVGDEGFAINLASNE
EALLELLIAIEKAGTTPGKDI SLALDCAASFSYNNVTGTYDGRHYEEQITALLNCDK
YPIDISDGLAEEYDQWALLTEVLGKQIVGDDLEVTNPDLILEGINSGLANSVLI
KPNQIGTLTETVYAIKLAQMAGYTTIISHRSETDTTIIADLAVAFNAGQIKTGLSR
SERVAKYRNLMEIEEELGSEAFITDSNYSFYEDSEE"
complement(5345..7318)
/gene="uvrB"
complement(5345..7318)
/gene="uvrB"
/codon_start=1
/transl_table=11
product="exinuclease ABC subunit B"
/protein_id="BAA99009.1"
/db_xref="GI:8979175"
/translation="MTFOLHAPFCQDOPEAIARLSAGVNRVQKSOVLIGTGGSGT
FTIANVAVNPLPTLHANKTLAQLQYOEPEFPNNNAVYFTSYDYQPEAYIAR
SDTIYERSLINDEIDKLRLSATRSITLBERDTLVSSVCYIGSPSTSMALVLE
VGKEYPRNLITLAQLVKHMQHQAIFIPORSAPRGSVIIFPAYESELARLEFLNDTL
TSIYSDPLTMIPIKSVSATLIPGSHVIYIPEAIREQAIRTIQBLEERMAFFDRPI
```

```

EKDRIFRTHTHDIEMKEIGFCCKGIENYSRHFTGAPPGAPPTCLLDYFPEDFLIIDE
SHQTLPOIRAMRGDSQKSLVEYGRFLPSADNRPDLTIEEAKYFRKVIYVSATPG
DTEVOESSGHVLIIRPTQIPDPMPETATQVQDVLLEIRLRLSKHKEKILIVSI
TKKLAEMAGLESELEIPAAVLSHGIEAERTQILTLDRSGVIDVLGVNLLREGLDL
PEVSLVAILDADKGGFLRSTSSLIQFCGGAARNKGVIFVADQKTRSIETRETER
RQIOLDYNKHNIVPKPIIKALFANPILQTSKDSKSPKSORPLSKEDLEEQIKYE
ALMQRAAEFRFNEAARYADMAQCKEQLLYLF"
complement(7327. .8361)
/gene="trps"
complement(7327. .8361)
/gene="trps"
/codon_start=1
/transl_table=1
/product="tryptophanyl tRNA synthetase"
/protein_id="BAA99010.1"
/db_xref="GI:8979176"
/translation="MNKKRVLTGDRPTOKLHLGHVGSIKNRLQNSPEYDCFFII
ADHLTKTKKEVLDVNDHIEVSLGIDPTKSIYIYQSASPEIYVHLHPS
ALISINRMVGIPLSKDMARNAIEEGSLGYIGYPILOSADILLAKAQFVPGKNE
HIVETLRDIARNRLXQVFPPEVLQGLTSLVIGDOGKMSKANNAYILSDSA
TITERKRMVTDPNRIATTPGRVGNPLFIYVHDFNPHKDEVEFRKARYOGCIKI
EVKARLAEELIHFLEKPIKERSEFLSKPLALQNVLEDGTHKMRVAKVMEVHDKFG
FSHKWRSLLK"
complement(8370. .8924)
/gene="CPj0803"
complement(8370. .8924)
/gene="CPj0803"
/codon_start=1
/transl_table=1
/product="CT584 hypothetical protein"
/protein_id="BAA99011.1"
/db_xref="GI:8979177"
/translation="MAAKTKTLEDNVLLLEGNLKRIEATPIGYTTFREFQNVFN
CANGQOEIAFFEMLINGLTQELAPQOQAHLIAEFMMPIRVARDIHERGEFIN
FITSMDLQQPERCIFRLARLVQDQFELMTDVQNTCHLIRLLARLLEAKNPVGK
NLOEQEETISLKNHFDLTKALO"
complement(8928. .9680)
/gene="gp60"
complement(8928. .9680)
/gene="gp60"
/codon_start=1
/transl_table=1
/product="CHLTR plasmid paralog"
/protein_id="BAA99012.1"
/db_xref="GI:8979178"
/translation="MGNLKTLLSREKKNPTKMEALARKWEGDPSPLAVRLSNPTL
SSKEKQRLHLLHYNREQIEPDLTOLCTLSAPVKOIHQSOLLGERITKVRDL
KRYRGCASSLLITYGNROTPTNFVYIELFTLLPELPKTEMEKMPQAVYVTLASRQ
GQEKREIIRNFRGKSELDRKIRKPEPLVETDCKRTSPVKQLAWLRTGSOILFK
CISLSDDQIIEKLKLEKVKVSNLPDTKV"
complement(9685. .10452)
/gene="minb"
complement(9685. .10452)
/gene="minb"
/codon_start=1
/transl_table=1
/product="chromosome partitioning ATPase"
/protein_id="BAA99013.1"
/db_xref="GI:8979179"
/translation="MKTIAYNSFKGTAKTSTTLHLGAALQYHOARVLLIDPDAQAN
LTSGLGDCYDCLAVVLOGEIEQIEVIRPQDLOLDLPADTWLERIEVSGNLAAD
RYSHERLQYVLGSDQVDYVVIDTTPSICMLTSESALIAADYALICATPEPYSVKGLE
RLAGFTQGISARHPPLTLGLVALSFWNCRCGNKNSAFELHKHTFPGKLLNTKIRDIIV
SFAAHGKVPFATSPSARASEDYENLTKEILLIRDI"
complement(10485. .12392)
/gene="thrs"
complement(10485. .12392)
/gene="thrs"
/codon_start=1
/transl_table=1
/product="threonyl tRNA synthetase"
/protein_id="BAA99014.1"
/db_xref="GI:8979180"

```

```

/translation="MIQVTCQKNYEVLEGTAAELAKQLKNSHQFIGVLINERPRDL
Query Match 100.0%; Score 1086; DB 1; Length 325865;
Best Local Similarity 100.0%; Pred. No. 2.7e-244;
Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGCAAGTACCTCAAAACACAGAAAGATATTCTCTGATTGTAAAGGGAATTCATTG 60
DB 295293 TTGCAAGTACCTCAAAACACAGAAAGATATTCTCTGATTGTAAAGGGAATTCATTG 295352
QY 61 TAACAGACCTTGGTCGAGCATCCATCTTAATTTTGAAGTGGTCTTACCAAGTTCATCAAA 120
DB 295353 TAACAGACCTTGGTCGAGCATCCATCTTAATTTTGAAGTGGTCTTACCAAGTTCATCAAA 295412
QY 121 ATGCTTATTAATAAAGTCTCTTGTGGAGTTCATCTATTGAAGGGAAGACGAATAAT 180
DB 295413 ATGCTTATTAATAAAGTCTCTTGTGGAGTTCATCTATTGAAGGGAAGACGAATAAT 295472
QY 181 CTGATCCAATTTCTAAAGGGAGAACACGTACGTATTCAATGAAGCGTGTCCCAAGTTCCTG 240
DB 295473 CTGATCCAATTTCTAAAGGGAGAACACGTACGTATTCAATGAAGCGTGTCCCAAGTTCCTG 295532
QY 241 AGGAGAACTGATGAAGAACGCTCCATATGCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300
DB 295533 AGGAGAACTGATGAAGAACGCTCCATATGCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 295592
QY 301 GCTTTCACAGTAAACCAACGAGCGGTACAAATAGGATGAGATGCTTTCTAGGCTTTG 360
DB 295593 GCTTTCACAGTAAACCAACGAGCGGTACAAATAGGATGAGATGCTTTCTAGGCTTTG 295652
QY 361 TATCGGTTGTTTTTACTAGGAAGAGCTTGTAATCTTTTCTATGACTTTTATGACTTTAGTGAAGTCT 420
DB 295653 TATCGGTTGTTTTTACTAGGAAGAGCTTGTAATCTTTTCTATGACTTTTATGACTTTAGTGAAGTCT 295712
QY 421 TCCCATGTTTATGAANAATTCAGANAATGAGGATGTAGAAGCCAAAGCACGTGCCTCAA 480
DB 295713 TCCCATGTTTATGAANAATTCAGANAATGAGGATGTAGAAGCCAAAGCACGTGCCTCAA 295772
QY 481 AAGGTCTATGATGAGATCAGGAGTAAAGGAGAAATTAATCTTTTATGAGAGTTCGCAATTAC 540
DB 295773 AAGGTCTATGATGAGATCAGGAGTAAAGGAGAAATTAATCTTTTATGAGAGTTCGCAATTAC 295832
QY 541 AGAGCAATGGGAAGTAAGCGTCTGCAACAGACAGCAAGTATTGATGGGAGCTACAAA 600
DB 295833 AGAGCAATGGGAAGTAAGCGTCTGCAACAGACAGCAAGTATTGATGGGAGCTACAAA 295892
QY 601 TCCTGAACCGGGATAATAGGANAATGTCTTCTAATTCGAAGAATTTAATGACTGCTTG 660
DB 295893 TCCTGAACCGGGATAATAGGANAATGTCTTCTAATTCGAAGAATTTAATGACTGCTTG 295952
QY 661 ATGCTCTAAGAAATCAGAAAGGAGACCGGAAAGCTTCTTAAGTAATCAAGAAAAATCTTGG 720
DB 295953 ATGCTCTAAGAAATCAGAAAGGAGACCGGAAAGCTTCTTAAGTAATCAAGAAAAATCTTGG 296012
QY 721 ACAATAGCTTTTCCATAGAGAAAACCGTGGAGAAAGAGTTCGATACCATTAATTCGGATT 780
DB 296013 ACAATAGCTTTTCCATAGAGAAAACCGTGGAGAAAGAGTTCGATACCATTAATTCGGATT 296072
QY 781 AAAAAGGAGTGTCTACAAACGCCATATCAGTGGAGCAACCTTCCCTTTGATTCTTTGA 840
DB 296073 AAAAAGGAGTGTCTACAAACGCCATATCAGTGGAGCAACCTTCCCTTTGATTCTTTGA 296132
QY 841 TAATCACAGAGATGACAAAGAGTAAAGTTTGAANAATTTCTAAATACANAATGCATTT 900
DB 296133 TAATCACAGAGATGACAAAGAGTAAAGTTTGAANAATTTCTAAATACANAATGCATTT 296192
QY 901 GTGTCGAGCTAAAGCTGCTTCTCTTTATTTTCTCTTTTGTAGATGATTCGGTAGTAGGA 960
DB 296193 GTGTCGAGCTAAAGCTGCTTCTCTTTATTTTCTCTTTTGTAGATGATTCGGTAGTAGGA 296252
QY 961 ATATGGGTCGAGTATGACAGTGGAGTGAATTTGGCTGTCTTGAAGAACAGGAAAGTGA 1020
DB 296253 ATATGGGTCGAGTATGACAGTGGAGTGAATTTGGCTGTCTTGAAGAACAGGAAAGTGA 296312

```







QY	658	TTGATGCTCTAAGAAATCAGAAAGGAGACCGGAAAGCTTCTAAGTAATCAAGAAATAATCT	717
Db	104552	GAGATTAATAAATAAAGATCAATCAACACAGAGTTTCTCTTTTAAAAAATAAGTAAAT	104493
QY	718	TGGACATACCTTTTCCATAGAGAAACCGTGAGGAAGGAGTCGATACCAATAATCCG	777
Db	104492	TGACATCTTTTAGCTACACTAAGAAAAAAGAGAGACACTCAAAATTTACAAAAATAA	104433
QY	778	ATTTAAAAAGGGAGTCTTTACACAGCCATATCAGTCGACCACTTCCCTTTGATTCTT	837
Db	104432	AGGTGAAAAAGGAAACTTAACTTAACCTGACACACAGTAATACACAGATACACAGACAT	104373
QY	838	TCATATGACAGAGATGACAAAAGAGTAAGTTTGTGAAAAA	879
Db	104372	TATGAACAACCTGTATACCAACAATTTGGAATCTAGAAAA	104331
RESULT 8			
AFI135186/c			
LOCUS	AFI135186	1642 bp	mRNA
DEFINITION	Caenorhabditis elegans tryptophan hydroxylase (tph-1) mRNA, INV 13-MAY-1999 complete cds.		
ACCESSION	AFI135186		
VERSION	AFI135186.1	GI:4809152	
KEYWORDS	Caenorhabditis elegans.		
ORGANISM	Caenorhabditis elegans.		
REFERENCE	Size, J. Y. and Ruvkun, G.		
AUTHORS	tph-1 encodes a C. elegans tryptophan hydroxylase		
TITLE	Unpublished		
REFERENCE	Size, J. Y. and Ruvkun, G.		
AUTHORS	Submitted (16-MAR-1999) Department of Molecular Biology/Department		
TITLE	of Genetics, Massachusetts General Hospital/Harvard Medical School,		
JOURNAL	Boston, MA 02114, USA		
FEATURES	Location/Qualifiers		
source	1..1642		
gene	/organism="Caenorhabditis elegans"		
CDS	/db_xref="taxon:6239"		
	/chromosome="II"		
	1..1642		
	/gene="tph-1"		
	44..1642		
	/note="tph-1"		
	/codon_start=1		
	/product="tryptophan hydroxylase"		
	/protein_id="AAD30115.1"		
	/db_xref="GI:4809153"		
	/translations="MDSLFQMASAMKFOYYSKKAAGTKMTSMNSVMSDNRMDEPKRFR		
	RRSGLGIPVPVEDVKQFLFTPTVRREASIREGDEEGVQILTIIVKSSRVSDIS		
	KMIANLPDHPRIKHELRDSODGSKTMDVLELPHYKQKQAMDLMLRLGLDVEV		
	SSTIRPTAIKEQYTEPGSDATTGSEWPKSIYDLIDICAKRVIMYGAGLDHPGFKD		
	TEYQRMRMFAELALNKHGEPIPREYTSERKTWGIYYKLRELHKHACKQFLDN		
	PELLERHCYGSNNIPQLEDICKFLKAKTGFVRPVPVAGYLSARDLAGLAYRVFFCTQ		
	YVRHHDPFTYPPEDTVVHELMGMALFADPDFQSOEIGLASLASEEDLKKLATLY		
	FFSIEFLGSLDDAASPVKENGSNHEREKVVYAGLLSAGELQHAVEGSATIIREFDPD		
	RVVFEQLITTFQSAFYFRNFEEAQOKLRMTNNMKRPPIVRNPNYTESVLENNRS		
	SIMLVNLSRSDINLLAGALHYIL"		
BASE COUNT	479 a	347 c	351 g 465 t
ORIGIN			
Query Match	4.3%;	Score 46.6;	DB 3; Length 1642;
Best Local Similarity	45.5%;	Pred. No. 1.5;	
Matches	249; Conservative	0; Mismatches	289; Indels 9; Gaps 2;
QY	113	TCATCAAAATGCTTATTCAAAATAAGTCTCTTGTGGAGTTGATGATTGAAGGAAGA	172



```

cDNA EST yk273b10.5 comes from this gene"
/codon_start=1
/protein_id="CAA91284.1"
/db_xref="GI:3878361"
/db_xref="SPTREMBL:Q21361"
/translation="MDVDSALSPSSRLSVVCSASAETSSSSSSSNFSGSPPEERR
NSVNESVTKDEHYERRRRNDASRRSRKRRQNDLAMEEKIMLLSENKRLSQIGT
TIPQSVTEPTSLIIPQVAKNLFAPAGTASLQASSMLTLPLOASHIPSMQLQCO
LPTQITQSVTASTQPAQSTASASSLFSSSSSAFHFPSPESAQOQSPSSVILVIER
RSPDSTVNMFPQLOQSGSVIQIQGAPAGSTPPQVIAQVQOQSPSLLSALLSQRRP
SPVTPQSRTEHIGLQNSPRHTGNKDCESVSSASFSFSSHSDHNSNKSPOYVD
RRRNNEAKRANRRRAVFEYRSRRVOLLEGENEDLTQIETLKAETIAHFKSVLAOR
ASVVTALHP"
complement(join(18851..19061,19149..19267,19316..19421,
19473..19561,19609..19713,19799..19939,19989..20176,
20245..20482,20530..20599,20934..21022))
/gene="K08F8.3"
complement(join(18851..19061,19149..19267,19316..19421,
19473..19561,19609..19713,19799..19939,19989..20176,
20245..20482,20530..20599,20934..21022))
/gene="K08F8.3"
/notes="contains similarity to Pfam domain: PF00852
(Fucosyl transferase), Score=875.1, E-value=6.9e-260, N=1
cDNA EST yk78d11.3 comes from this gene
cDNA EST yk340d5.3 comes from this gene
cDNA EST yk340d5.5 comes from this gene"
/codon_start=1
/protein_id="CAA91285.1"
/db_xref="GI:3878362"
/db_xref="SPTREMBL:Q21362"
/translation="MKKQNTPRVGYATSCRMVLGVDRFLRLKWLMPACCTYLL
VIVAPTSQEQDKKEGEIJSNDHELDVPILOKEELKQOQSPSEENVPKRRTNFN
PVKEPFDVEVLTSQDKLEERTATVPIQKRLILLSNAGHSQNDLQGGPDWCFE
TQVRAAPADADAVLAHMDNDVPKPNQVYVFSQSPANSFIQIPRDIINMILGR
HDTPASPGTYTVKLGAQRKGTQGVQANLVNGKAKGAAMFVSHQCTSKRDEPVKLL
QKHLIDIVGGCGPKKARGSKCDTMDLTDYHFYTFENSICEDYVTEKWLKSGYQN
TLPLVLKRLPEPEVFNKPSFIADDFKSKEMGDYLNLMNNKTATVFEYFEWHDYK
VVFDDGSHHDVLERPWGFCQVCRMAWTEPRQKVLIPNWDAYWRQTECKDGLTVDSIPL
D"
complement(join(24990..25164,25210..25439,25484..25610,
25786..25921,25972..26168,26281..26443,26496..26688,
27097..27213,27309..27344))
/gene="K08F8.4"
complement(join(24990..25164,25210..25439,25484..25610,
25786..25921,25972..26168,26281..26443,26496..26688,
27097..27213,27309..27344))
/gene="K08F8.4"
/notes="contains similarity to pfam domain: PF00351
(Riboflavin-dependent aromatic amino acid hydroxylase),
Score=800.0, E-value=1.6e-252, N=1; PF01842 (ACT domain),
Score=89.6, E-value=2.2e-11, N=1
cDNA EST yk432h11.3 comes from this gene
cDNA EST yk432h11.5 comes from this gene
cDNA EST yk349f13.5 comes from this gene"
/codon_start=1
/protein_id="CAA91286.1"
/db_xref="GI:3979833"
/db_xref="SWISS-PROT:P90925"
/translation="MPPAGDDLDLKYAMESYVADNVADNIGKTTIVFLREKAGALA
ETLKLQADVNLSHIESRPSKTEGCEYVLEFAEADHRKIEGVIEHFQOKAEKKV
LVQWNTNKNQKDSVPWFQKINDIDQFANRLSYGAELDADHPGKDMTYKERRFK
FADIAFNKGGDKIPTITTYDEIATVTRNLTVMYPRKNAQOEFNIFPLLLQKNCQ
FGPDRIPQLOQVSDFLKDCGTGTRFVAGLLSSRDLFAGLAFRVFSTQYIRHHSAPK
YTPEDICHELLHVLFLADVEFAQFSQIGLASLAPDDVIEKLATLYWTFTEFGIC
QODGKKAYGAGLSDSFELOALSDKPEVDFDPVAVCCVTKYPIITEYQPKYFLAES
ASAKNKLKSWAATINRPFQIRYNAYTORVEILDVKVAALQRLARDIRSDISTLEEALGK
VNNLKNK"
join(29693..29728,30107..30376,30419..30588,30762..30896,
30946..31051,31100..31190,31450..31580)
/gene="K08F8.5b"
join(29693..29728,30107..30376,30419..30588,30762..30896,
30946..31051,31100..31190,31450..31580)

```

```

/gene="K08F8.5b"
/codon_start=1
/protein_id="CAA91288.1"
/db_xref="GI:3878365"
/db_xref="SPTREMBL:P90926"
/translation="MGRANDSGVVKMFVLDKTIYSGADYGGSTOLLILDHIEKSKS
KYDMVNEKKASQLTIEGKYEMVRCYSASMTWPEHGLLGVNTEKLVAHFVDNEMOI
KTTKLEULLVQKLTAGRDATKEKEKSTGSKSLKLLTVNDWEDENFGSAMPNSKN
VMSIKNTPSAGIDQFVALRYRGKPTMGRANSGRIEASFVDAKREFTCATVGSLO
LIISLPTTYGDDYVWMTYTOADYDKQYVSHLSYICPAIVPVDCKFFLGQVNMKT
ECATVALNCSVTDLDGSPVASIIVLCRKEMATEMLI"
join(29858..29977,30107..30376,30419..30588,30762..30896,
30946..31051,31100..31190,31450..31580)
/gene="K08F8.5a"
join(29858..29977,30107..30376,30419..30588,30762..30896,
30946..31051,31100..31190,31450..31580)
/gene="K08F8.5a"
/codon_start=1
/protein_id="CAA91287.1"
/db_xref="GI:3878364"
/db_xref="SPTREMBL:Q21365"
/translation="MESHEVLIPEWPNILQIGILPTVDMCSGKLDRAKFLQICMFV
LDDKIYSGADYGGSTQILLIHDHIEKSFHYDMWNFKKASQLTIEGKYEMVRCYSAS

```

Query Match 4.2% Score 45.6; DB 3; Length 41246;  
Best Local Similarity 59.1%; Pred. No. 1.8;  
Matches 78; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 191 TCTAAGGAGCAACACGTCATCAATGAAAGCGTGTCACAGCATATGAAGCTCTCCAAATGAATC 250  
DB 25306 TCAAAATCAACATCTCCGCTTATCACTCAACGATATGAAGCTCTCCAAATGAATC 25365  
QY 251 ATAAAGACGGTCATATGCTTTCTTCCTCATGTTTCAATAAGTCGCTTTCAACA 310  
DB 25366 AAAATCGGCTCGTAGCTTTTCTCCCATCTTGTGCACAGATTCCAAAATTCGATT 25425  
QY 311 GTAAACCAAAAG 322  
DB 25426 GTGAACCAAGTAG 25437

RESULT 11  
AC121238/c  
LOCUS  
DEFINITION  
AC121238  
AC121238.8 GI:22218484  
HTG: HTGS\_PHASE2; HTGS\_DRAFT.  
KEYWORDS  
SOURCE  
ORGANISM  
Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
Medicago.  
REFERENCE  
1 (bases 1 to 109206)  
Shauli, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,  
Cook, D., Kim, D. and Roe, B.A.  
Medicago truncatula BAC Clone mth2-34p9  
Unpublished  
2 (bases 1 to 109206)  
Shauli, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,  
Cook, D., Kim, D. and Roe, B.A.  
Direct Submission  
Submitted (16-MAY-2002) Department of Chemistry And Biochemistry,  
the University of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
3 (bases 1 to 109206)  
Shauli, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,  
Cook, D., Kim, D. and Roe, B.A.  
Direct Submission  
Submitted (23-AUG-2002) Department of Chemistry And Biochemistry,  
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,



Db	863	BAYTNNNNRMYGAYADDYAYTMSDTCDAWKKWDATKMNATYTRNGTAWRNNNN	922
QY	864	GTAAGTTTGTGAAAAATCTAAATAGAAATGCAITTTGTCTCAGCTGAAGCTGCTCT	923
Db	923	NMTMKTYXBHAAHNNNNNGKMCATHTFWCKAKTKTTCGWNMCNTTTCRKTNNCWT	982
QY	924	CTTATTTTCTCTTTTGTAGATGATTCGGTAGTAGGAATATGGGGTCGATGACACGTT	983
Db	983	WMTTTRTWTAAATRWKTNATSMTRCATGKNNNTWTGKTRWTAIRMTATRWKAWK	1042
QY	984	CGAGTGTGCTGCTCTCGAAGAACAGGGAAGTATCTCTCAGCTTTTACAGCAATCT	1043
Db	1043	VNATGSWNTNSYARWYKTRAYKGYWYNACAWRWGKATCYMTDNATCATATSWMATH	1102
QY	1044	TCAGAATATATCTTTGGGTGCCA	1064
Db	1103	KYNWCMCKNNNNNNNTMMR	1123
RESULT	13		
AC024941			
LOCUS		172246 bp	DNA linear PRI 11-JUL-2001
DEFINITION		Human sapiens 12 BAC RP11-900F13 (Roswell Park Cancer Institute Human BAC Library) complete sequence.	
ACCESSION		AC024941	
VERSION		AC024941.30	GI:14669928
KEYWORDS		HTG.	
SOURCE		Homo sapiens.	
ORGANISM		Homo sapiens	
REFERENCE			
AUTHORS			
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
		1 (Bases 1 to 172246)	
		Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-ozman,F.R., Allen,C., Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,M., Bryant,N.P., Buhay,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,F.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homsf,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,B., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratoch,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,B., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H., Lozada,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Maja,P., Massey,E., Mawhiney,E., McLeod,M.P., Martindale,A., Martinez,B., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Meischer,S., Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,I., Monabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Prims,E., Pui,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojokan,I., Rolfe,M., Ruiz,S., Savery,C., Scherer,S., Scott,G., Shen,H., Shim,C., Shoshitari,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R. and Gibbs,R.	

TITLE	Direct Submission	
REFERENCE	2 (bases 1 to 172246)	
AUTHORS	Worley,K.C.	
TITLE	Direct Submission	
JOURNAL	Submitted (03-MAR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	
REFERENCE	3 (bases 1 to 172246)	
AUTHORS	Worley,K.C.	
TITLE	Direct Submission	
JOURNAL	Submitted (11-JUL-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza Houston TX 77030, USA	
COMMENT	On Jul 11, 2001 this sequence version replaced gi:13492990. INFORMATION: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a> or email <a href="mailto:gc-help@bcm.tmc.edu">gc-help@bcm.tmc.edu</a>	
	CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.	
	ANNOTATION OF FEATURES:	
	STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.	
	Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.	
	Genes and region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.	
	SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.	
	QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <a href="http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html">http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html</a> .	
	QUALSTAT-REPORT-----	
	----- Summary Statistics -----	
	Contig length: 172246	
	Phrap values in estimate: 171130	
	Average error rate (BCM-Phrap estimate): 2.13165e-05	
	Fraction of phrap values less than 40 : 13	
	Number of consensus changing edits: 0	
	Number of N's in consensus :	
	----- Consensus changing edits -----	
	Position Original+Context Edited+Context	
	67598 tcttagacag(n)aatgacaata tcttagacag(c)aatgacaata	
	99512 cctcttggaaa(n)tattatttg cctcttggaaa(c)tattatttg	
	103635 ggcattgggtg(n)acaggccccc ggcattgggtg(c)acaggccccc	
	103636 gcattgggtg(n)acaggccccc gcattgggtg(c)acaggccccc	
	103660 aaagactgga(a)ccagattgg aaagactgga(a)ccagattgg	
	106704 tatattacct(t)taaaaattt tatattacct(t)taaaaattt	
	106705 atattacctt(t)taaaaattt atattacctt(t)taaaaattt	





```
* 4054 4783: contig of 730 bp in length
* 4784 gap of unknown length
* 4884 5571: contig of 688 bp in length
* 5572 5671: gap of unknown length
* 5672 6376: contig of 705 bp in length
* 6377 6476: gap of unknown length
* 6477 7207: contig of 731 bp in length
* 7208 7307: gap of unknown length
* 7309 7613: contig of 306 bp in length
* 7614 7713: gap of unknown length
* 7715 8902: contig of 1189 bp in length
* 8903 9757: gap of unknown length
* 9758 9857: gap of unknown length
* 9858 10343: contig of 486 bp in length
* 10344 10443: gap of unknown length
* 10444 11201: contig of 758 bp in length
* 11202 12335: contig of 1034 bp in length
* 12336 12435: gap of unknown length
* 12436 13294: contig of 859 bp in length
* 13295 13394: gap of unknown length
* 13395 14096: contig of 702 bp in length
* 14097 15252: contig of 1056 bp in length
* 15253 16686: gap of unknown length
* 16687 17422: contig of 1314 bp in length
* 17423 17522: gap of unknown length
* 17523 18416: contig of 636 bp in length
* 18417 18516: gap of unknown length
* 18517 19058: contig of 894 bp in length
* 19059 20107: contig of 542 bp in length
* 20108 20207: gap of unknown length
* 20209 21392: contig of 949 bp in length
* 21393 22510: gap of unknown length
* 22511 22610: contig of 1185 bp in length
* 22611 23648: gap of unknown length
* 23649 23748: contig of 1018 bp in length
* 23749 24975: contig of 1038 bp in length
* 24976 25075: contig of 1227 bp in length
* 25076 26075: gap of unknown length
* 26076 26175: contig of 1000 bp in length
* 26176 27416: gap of unknown length
* 27417 27516: contig of 1241 bp in length
* 27517 28528: gap of unknown length
* 28529 28628: contig of 1013 bp in length
* 28630 29583: gap of unknown length
* 29584 30741: contig of 954 bp in length
* 30742 30841: gap of unknown length
* 30842 32588: contig of 1058 bp in length
* 32589 33083: gap of unknown length
* 33084 33154: gap of 1148 bp in length
* 33155 33254: gap of unknown length
* 33255 34457: contig of 1065 bp in length
* 34458 34557: gap of unknown length
* 34558 35986: contig of 1203 bp in length
* 35987 36086: gap of unknown length
* 36087 37925: contig of 1429 bp in length
* 37926 38025: gap of unknown length
* 38026 39294: contig of 1839 bp in length
* 39295 40332: gap of unknown length
* 40333 40532: contig of 1269 bp in length
* 40533 42129: gap of unknown length
* 42130 42229: contig of 1138 bp in length
* 42230 43855: contig of 1497 bp in length
* 43856 43955: gap of unknown length
* 43956 45742: contig of 1626 bp in length
* 45743 gap of unknown length
* 45744 45843: contig of 1787 bp in length
* 45844 gap of unknown length
* 45845 46800: contig of 730 bp in length
* 46801 46899: gap of unknown length
* 46900 48310: contig of 688 bp in length
* 48311 48409: gap of unknown length
* 48410 48769: contig of 705 bp in length
* 48770 49669: gap of unknown length
* 49670 51672: contig of 731 bp in length
* 51673 51772: gap of unknown length
* 51773 53070: contig of 731 bp in length
* 53071 53170: gap of unknown length
* 53171 55690: contig of 306 bp in length
* 55691 55990: gap of unknown length
* 55991 58340: contig of 1189 bp in length
* 58341 58440: gap of unknown length
* 58441 60742: contig of 755 bp in length
* 60743 60842: gap of unknown length
* 60843 62552: contig of 486 bp in length
* 62553 62653: gap of unknown length
* 62654 65164: contig of 758 bp in length
* 65165 65263: gap of unknown length
* 65264 66984: contig of 1034 bp in length
* 66985 67085: gap of unknown length
* 67086 68551: contig of 859 bp in length
* 68552 70725: gap of unknown length
* 70726 70825: contig of 702 bp in length
* 70826 72615: gap of unknown length
* 72616 72715: contig of 1056 bp in length
* 72716 74765: gap of unknown length
* 74766 77074: contig of 1314 bp in length
* 77075 77174: gap of unknown length
* 77175 79340: contig of 636 bp in length
* 79341 80990: gap of unknown length
* 80991 81090: contig of 894 bp in length
* 81091 82975: gap of unknown length
* 82976 83075: contig of 542 bp in length
* 83076 84926: gap of unknown length
* 84927 85026: contig of 949 bp in length
* 85027 87534: gap of unknown length
* 87535 87634: contig of 1185 bp in length
* 87635 90562: gap of unknown length
* 90563 90662: contig of 1018 bp in length
* 90663 92287: gap of unknown length
* 92288 92386: contig of 1038 bp in length
* 92387 95352: gap of unknown length
* 95353 95451: contig of 1227 bp in length
* 95452 97621: gap of unknown length
* 97622 97720: contig of 1000 bp in length
* 97721 100394: gap of unknown length
* 100395 100494: contig of 954 bp in length
* 100495 102562: gap of unknown length
* 102563 102662: contig of 1058 bp in length
* 102663 105607: gap of unknown length
* 105608 105707: gap of 1148 bp in length
* 105708 108535: contig of 1065 bp in length
* 108536 108635: gap of unknown length
* 108636 111168: contig of 1203 bp in length
* 111169 111268: gap of unknown length
* 111269 114040: contig of 1429 bp in length
* 114041 114140: gap of unknown length
* 114141 117073: contig of 1839 bp in length
* 117074 117173: gap of unknown length
* 117174 119836: contig of 1269 bp in length
* 119837 119937: gap of unknown length
* 119938 123089: contig of 1138 bp in length
* 123090 123189: gap of unknown length
* 123190 125696: contig of 1497 bp in length
* 125697 125797: gap of unknown length
* 125798 128622: contig of 1626 bp in length
* 128623 128722: gap of unknown length
```







QY 750 AGGAAAGGAGTCATACCATCAATAATTCGATTAAAAA 786

DB 222544 ATAAAAATTTTAAATTAATAACAAAAATAAAAAA 222508

## RESULT 18

AC008008

LOCUS

AC008008 Homo sapiens Xp22 PAC PC16-102 (Roswell Park Cancer Institute

DEFINITION Human PAC Library) complete sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AC008008 167257 bp DNA linear PRI 16-MAY-2002

HMG. AC008008.2 GI:5656683

Homo sapiens.

Homo sapiens.

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 167257)

Muzny, D., Arenson, A. D., Bouck, J., Brundage, E., Bunac, C., Chen, Z.,

Di, W., Ding, Y., Dugan, S., Durbin, J., Forcum, J., Garcia, C.,

Gorell, J. H., Gorell, B., Hernandez, J., Liu, W., Logan, O., Lu, J.,

Kondejewski, N., Leal, B., Lichtarg, O., Liu, W., Perez, L.,

Martinez, C., Osval, G., Pampell, L. R., Parish, B. J., Perez, L.,

Rashid, N. D., Rives, C., Scherer, S. E., Shen, H., Simon, M., Vo, Q.,

Williamson, A., Worley, K. C., Yu, W., Zhou, X., Nelson, D. and

Gibbs, R. A.

Direct Submission

Unpublished

2 (bases 1 to 167257)

Worley, K. C.

Direct Submission

Submitted (10-JUL-1999) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 167257)

Worley, K. C.

Direct Submission

Submitted (31-JUL-1999) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

5 (bases 1 to 167257)

Worley, K. C.

Direct Submission

Submitted (09-OCT-1999) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

6 (bases 1 to 167257)

Worley, K. C.

Direct Submission

Submitted (16-MAY-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Jul 30, 1999 this sequence version replaced gi:5441911.

INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email

qc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the

entire insert of this clone. Overlapping regions of clones are only

sequenced and submitted once, so the sequence for the remainder of

the insert may be found in the record for the adjacent clones.

Overlapping clones are noted at the beginning and end of the

Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches

of a local database that includes entries from dbSTS, GDB, and

local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

## QUALSTAT-REPORT-----

## ----- Summary Statistics -----

Contig lengths: 167258  
Phrap values in estimate: 166248  
Average error rate (BOM-Phrap estimate): 0.000259222  
Fraction of Phrap values less than 40: 0.0376546  
Number of consensus changing edits: 28  
Number of N's in consensus: 0

## ----- Consensus changing edits -----

Position	Original-Context	Edited-Context
11124	ttgtgagac(n)attnnccatg	ttgtgagac(g)attgacatg
11128	gatgacatt(n)ncatgaaat	gatgacatt(g)acatgaaat
11129	atgacatt(n)ncatgaaat	atgacatt(g)acatgaaat
11130	tgacatt(n)ncatgaaat	tgacatt(g)acatgaaat
11160	tgattactag(n)gataactact	tgattactag(t)gataactact
32920	aaaggctcag(n)agatgcagt	aaaggctcag(g)agatgcagt
32998	actgcatct(n)tcacatccag	actgcatct(c)tcacatccag
33125	tgagcattt(n)caggtatatt	tgagcattt(a)caggtatatt
64965	ttttcttcc(n)tttttttt	ttttcttcc(t)tttttttt
70901	tcttacatca(n)ataaataag	tcttacatca(a)ataaataag
72799	caccagatta(n)ttgaatgaa	caccagatta(a)ttgaatgaa
119977	aattaccaaa(n)gtggggggg	aattaccaaa(t)gtggggggg
119982	ccaaagtgg(n)ggggncagg	ccaaagtgg(g)ggggncagg
119987	ngtgggggg(n)caggatttga	ngtgggggg(t)caggatttga
120157	caacacttag(n)gggttataaa	caacacttag(t)gggttataaa
120195	tggtgtctg(n)ggatggacta	tggtgtctg(t)ggatggacta
121880	aattcaccca(n)naaagttagc	aattcaccca(a)naaagttagc
121881	attcaccca(n)naaagttagc	attcaccca(a)naaagttagc
124786	agaaatgga(n)aaaaaat	agaaatgga(a)aaaaaat
141954	ataagtat(n)tgatttcatt	ataagtat(c)tgatttcatt
150222	ttattctgat(n)tgctctct	ttattctgat(t)tgctctct
150300	ctgaggacta(n)ctaaacacg	ctgaggacta(c)ctaaacacg
150332	aaataccctt(n)cataagacat	aaataccctt(c)cataagacat
150354	ccacacatg(n)gccatgtcag	ccacacatg(t)gccatgtcag
150594	cgttgggctc(n)tmaggagatg	cgttgggctc(g)tmaggagatg
160398	ttgggctctt(n)naggagatta	ttgggctctt(g)naggagatta
160400	ttgggctctt(n)naggagatta	ttgggctctt(g)naggagatta
160401	ttgggctctt(n)naggagatta	ttgggctctt(g)naggagatta

## ----- Distribution of Quality &lt; 40 Bases -----

Quality	Count
10001	1
9001	1
8001	1
7001	1

```

bases      600|          * * * * *
500|         * * * * *
400|         * * * * *
300|         * * * * *
200|         * * * * *
100|         * * * * *
0|          * * * * *
-----
          5  10  15  20  25  30  35  40
Phrap Value Range

-----
Version: 1.01 xfigo
Location/Qualifiers
1..167257
/db_xref="taxon:9606"
/chromosome="X"
/clone="RPC16-102"
/complement(1142..1426)
/rpt_family="Charlie5"
1531..1578
/rpt_family="L2"
2470..2525
/rpt_family="GA-rich"
3178..3310
/rpt_family="LTR16C"
/complement(4538..4571)
/rpt_family="L2"
4690..4908
/rpt_family="MER58A"
/complement(5206..5286)
/rpt_family="MIR"
5397
/rpt_family="MER20"
/complement(5401..5554)
/rpt_family="MIR"
7110..7186
/rpt_family="MER91C"
7144..7353
/standard_name="DXS9963"
/db_xref="dbSTS:64811"
/complement(8394..8628)
/rpt_family="AluSg"
/complement(9633..9771)
/rpt_family="MIR"
9907..11222
/rpt_family="L1MB8"
11244..11549
/rpt_family="AluJb"
/complement(11720..11846)
/rpt_family="L2"
12374
/complement(11924..12374)
/rpt_family="MER31B"
14349..14440
/rpt_family="L2"
/complement(14856..15171)
/rpt_family="AluJb"
/complement(15829..16052)
/rpt_family="L1MA7"

Query Match      4.1%; Score 44; DB 9; Length 167257;
Best Local Similarity 46.9%; Pred. No. 3.6;
Matches 137; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 578 AAGTATTGATGGGACGACAAATCCTCAACGCGGATAATAGGAAAAATGCTGTTCTAAT 637
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 71228 AAGTATACCTCAATATCTTGACTTTTAAAAATAAATAAATAGATGAGTCTCTC 71287

QY 638 TCGAAGAAATTAATGACTGCTTGATGGCTCAAGAAATCAGAAAGAGACCGAAAGCTTCT 697
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 71288 AAGAAAAATATCTATATTTTGTCCAGGAATGAATGCTATATACCATGAAGATGG 71347

```

```

QY 698 AAGTAATCAAGAAAAAATCTTGGACAATAGCTTTTCCATAGAGAAAAACCGTAGGAAGG 757
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 71348 AAGTTTTCTATAAGAGGCGCATTTATCGGACTCAATGACAATTTGAAAGCGCAAGTATTTA 71407
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 758 AGTCGATACCAATAATTCGGATTAAAAAGGAGTGCCTTACAACAGCCATATCAGTGGAG 817
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 71408 AAGGGAATTAATATTTCTCTGGCAAGATATGAATTACCAAAATTTGGCCACAGAAAGTAG 71467
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 818 CAACCTTCCTTTTGTGATTCATGACAGATGACAAAGATCAACAAAGAGTAAGT 869
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 71468 AAACATCAAAAAAGTAACACATAATCATAGATGATATTCAAAAAGTAATTT 71519
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 19
AC092662      198230 bp      DNA      linear      PRI 01-MAR-2002
LOCUS      Homo sapiens BAC clone RP11-510H11 from 2, complete sequence.
AC092662 AC026578
AC092662.2 GI:16303564
HTG.
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 198230)
AUTHORS      Sulston J.E. and Waterston R.
TITLE      Toward a complete human genome sequence
JOURNAL      Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE      95063792
PUBMED      9847074
REFERENCE
2 (bases 1 to 198230)
AUTHORS      Harris A., Abbott A., Dixon R., Dignan G. and Phillips A.
TITLE      The sequence of Homo sapiens BAC clone RP11-510H11
JOURNAL      Unpublished (2001)
REFERENCE
3 (bases 1 to 198230)
AUTHORS      Waterston R.H.
TITLE      Direct Submission
JOURNAL      Submitted (19-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
4 (bases 1 to 198230)
AUTHORS      Waterston R.H.
TITLE      Direct Submission
JOURNAL      Submitted (20-OCT-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
5 (bases 1 to 198230)
AUTHORS      Waterston R.H.
TITLE      Direct Submission
JOURNAL      Submitted (23-OCT-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
6 (bases 1 to 198230)
AUTHORS      Waterston R.
TITLE      Direct Submission
JOURNAL      Submitted (01-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 20, 2001 this sequence version replaced gi:14916246.
-----
Center: Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.edu
-----
Center project name: H_NH0510H11
Drafting Center: WIBR
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.



Matches 143; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

Qy 7 AAGTACTCAAAACAGAGATATTTCTTGATGTAAGGGGAATGATTTCAACAG 66  
 Db 102449 AAGATTTGAAAGAGCTCATTAATGATATTTTAAATTAATCAATTTGAACAGTT 102508

Qy 67 ACCTTGGTCGAGCATCCATCTTAATTTTGAAGTCAGTCTTACCAGTTCATCAAAATGTCT 126  
 Db 102509 ACCATATTTTATATCAATCAATATAATTTGAAGTCAGTTCATCAAAATCAAT 102568

Qy 127 TATTGAAATAAAGTCTCTGTTGGAGTGTATGATTAAGGGAAGCAATATCTGATC 186  
 Db 102569 TTGATCAATAGGCATCTAATGGAATGATGTTTGAACGAAATCAATTTTATAATA 102628

Qy 187 CAATTCTAAGGAGAACACGCTAGCTTATCAATGAAGAGCGTCTCCAGTTCCTGAGGAGA 246  
 Db 102629 CAAGATAAAGTGAAGCAGATATACAGATAAATGATATAAATGATAATTTGGGATT 102688

Qy 247 ACTGATAAAGCGGTCATATGCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 306  
 Db 102689 ATTTTCAGATAATTTCACTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 102748

Qy 307 AACAGTAA 314  
 Db 102749 TTTAGTAA 102756

RESULT 20  
 AX083744/c  
 LOCUS AX083744 1141 bp DNA linear PAT 28-FEB-2001  
 DEFINITION Sequence 22 from Patent WO0111061.  
 ACCESSION AX083744  
 VERSION AX083744.1 GI:13185472  
 KEYWORDS synthetic construct.  
 SOURCE synthetic construct.  
 ORGANISM Kunz, L. and Clemens, S.  
 REFERENCE 1 (bases 1 to 1141)  
 AUTHORS Regulation of embryonic transcription in plants  
 TITLE Patent: WO 011061-A 22 15-FEB-2001;  
 JOURNAL UNIVERSITY OF BRITISH COLUMBIA (CA)  
 FEATURES  
 source 1. 1141  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 promoter 1. 1141  
 /note="consensus sequence of A.L., L.a., and B.n. FAEL  
 promoters"  
 BASE COUNT 123 a 32 c 42 g 112 t 832 others  
 ORIGIN

Query Match 4.0%; Score 43.8; DB 6; Length 1141;  
 Best Local Similarity 11.6%; Pred. No. 6.9;  
 Matches 86; Conservative 248; Mismatches 405; Indels 1; Gaps 1;

Qy 348 TTGTAGGCTTGTATGCTGTTGTTTCTTACTAGGAGCTTGCTTCTTCTATGACTT 407  
 Db 1056 YTRSNANWSCATKBMWTKWYATYATYAWCAVRNNMWCATNGYAKSCATNNAMW 997

Qy 408 TAGTGAAGACTCTCCATGTTTATGAAATATGAGAAATGAGGATGAGAGCCCAAG 467  
 Db 996 YATTRWAAYAAKARWARAGNNRMWYGAAGKWKCMWAMTGBMWADTAGKMCNNNNN 937

Qy 468 GCAGTGCCTTAAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 527  
 Db 936 WTTDVRMAKAKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 877

Qy 528 GAGTTCGCATTAACAGAGGCAATGGGAAGTAAAGGCTGCTGCAACAGACACAAGATTGAT 587  
 Db 876 RTKYNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 817

Qy 588 GGGAGCTACAATCCTGAACGGGATAATAGGAAAAATGCTGTTTCTAATTCGAAGAATT 647

Db 816 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 757  
 Qy 648 TAATGACTGCTTGTATGCTTAAAGAAATCAGAAAGGAGAGACCCGAAAGCTTCTAAGATAACAA 707  
 Db 756 TWMTWYDMWMTTMBTNTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 697

Qy 708 GAAAAATCTTTGGACAATAGCTTTTCCATAGAGAAAAACCGTGAGGAAAGGAGTCGATACC 767  
 Db 696 CWNNTDARNTNTTVRRRMMNTKRWYSTTRHHYTGATNNNNNNNNNNNNNNNNNNNN 637

Qy 768 ATAATTTCCGATTAATAAGGAGCTCTTACACAGCATATCATGTCGACCAACCTTCCC 827  
 Db 636 CTCRTMTMTWTKWGDGTVRKVKWRDITCTIYVDWADSWWWYANWRCRDVTTTRNN 577

Qy 828 TTTGATCTTTTGTATGATGACAGATGAACAAAGAGTAAGTCTTTTGAAGAAATCTTAA 887  
 Db 576 TYKSYAHSYWSNNAMWYRRYSARNSSMARWTTTRNNWMSGBVMRWAGTMMWRHWN 517

Qy 888 TAGAATGCATTTGTCTGAGCTTAAAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 946  
 Db 516 NNNTDTRYWKKWRABTTTIVYDSMCNAKSMWRGNNNWRANKMMAANNDAGAMDHWTY 457

Qy 947 ATTCGCTAGTAAGTATGGGTGCGAGTATGACGCTTGGAGCTTGGAGTATGCTTCTGGAAG 1006  
 Db 456 MGNNTMMRRANKMMNNMWCRRAYCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 397

Qy 1007 AACAGGAAAGTATGCTCTCAGCTTTAGAGCAATCTTCAGAAATATATCTTTGGTCCAGG 1066  
 Db 396 AMYMRVAAMWYSRDNTTNDMMWMTSDWBWHWTVDYTMRAWNNNNNNNNNNNNNNNN 337

Qy 1067 GTTCTCTGCGAGTAGTCAC 1086  
 Db 336 WMDHNTHTCTYGNNTWGSAY 317

RESULT 21  
 AX344846/c  
 LOCUS AX344846 6089 bp DNA linear PAT 01-FEB-2002  
 DEFINITION Sequence 271 from Patent WO0200927.  
 ACCESSION AX344846  
 VERSION AX344846.1 GI:18492732  
 KEYWORDS synthetic construct.  
 SOURCE synthetic construct.  
 ORGANISM Olek, A., Piepenbrock, C. and Berlin, K.  
 REFERENCE 1  
 AUTHORS Diagnosis of diseases associated with development genes  
 TITLE Patent: WO 0200927-A 271 03-JAN-2002;  
 JOURNAL Epigenomics AG (DE)  
 FEATURES  
 source 1. 6089  
 Location/Qualifiers  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="chemically treated genomic DNA (Homo sapiens)"  
 BASE COUNT 1262 a 219 c 1585 g 3023 t  
 ORIGIN

Query Match 4.0%; Score 43.4; DB 6; Length 6089;  
 Best Local Similarity 45.8%; Pred. No. 7.1;  
 Matches 149; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

Qy 621 AAAAAATGTTTCTTAATTCGAAGATTTTAATGACTGCTTGTATGCTTGAAGAAATCAGAAA 680  
 Db 1350 AAAAAAATAATATTAACACCAATAATATATACACATAAATAAATAAATAAATAAATAA 1291

Qy 681 GGAGCCGAAGCTTCTTAAGTAATCAAGAAAAATCTTTGGACAATAGCTTTTCCATAGAG 740  
 Db 1290 AACTCTCTCAACCAATCTCAAAAAATCAAAAAAATCTCTCAAAAAAATATCTCTAAAA 1231

Qy 741 AAAACCGCTGAGGAGGAGTCGATACCAATAATTCGATTAATAAAGGAGTCTCTACAA 800



```

Db 1230 AAAATCCTAAAAAATAAATTAATTAATTAACAAATAAAATTTTAAACCAAAAAAACA 1171
QY 801 CAGCATATCAGTGGAGCAACCTTCCCTTTTGATTTCTTTGATATGACAGAGAGTGAACA 860
Db 1170 AACACGAAAAATTCAAAATACCTAATAATTAATAACGTTAAAAAATAAATAATATACAAA 1111
QY 861 AGAGTAAGTTTTGAAAAATCTTAAAAATAGAAATGCAATTTGTCTCGAGCTAAAGCTTGCT 920
Db 1110 AAAAACAATAAAAAATCTTAAAAACAACAACAATATTAATAAAAAATAACAATATTACA 1051
QY 921 TCTCTTATTTTCTTTTCTAGAAAT 945
Db 1050 TAATCAAAATTTTATTTTCAAAAAAT 1026

RESULT 22
AP001943/c
LOCUS AP001943 88966 bp DNA linear HTG 09-AUG-2000
DEFINITION Homo sapiens chromosome 4 clone 2185P9 map 4q22-q24, *** SEQUENCING
IN PROGRESS **, 7 unordered pieces.
ACCESSION AP001943.1 GI:7678833
VERSION HTG; HTGS_PHASE1.
KEYWORDS Homo sapiens DNA, clone:2185P9.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE
AUTHORS Tsai,S.F.
TITLE Direct Submission
JOURNAL Submitted (28-APR-2000) Shih-Peng Tsai, National Yang-Ming
University, Institute of Genetics, 155 Li-Rong St. Section 2,
Peitou, Taipei, Taiwan 11221, Republic of China
(E-mail: ymptsai@ym.edu.tw, URL: http://genome.ym.edu.tw/,
Tel: 886-2-28267043, Fax: 886-2-28264930)
COMMENT These sequences are draft human sequences, not finished sequences.
of the same clone are represented as 100 N.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved
* 1 26620: contig of 26620 bp in length
* 2 26720: gap of 100 bp
* 3 26721 31716: contig of 4951 bp in length
* 4 31672 31717: gap of 100 bp
* 5 31772 35904: contig of 4133 bp in length
* 6 35905 36004: gap of 100 bp
* 7 36005 44689: contig of 8685 bp in length
* 8 44690 44789: gap of 100 bp
* 9 44790 61582: contig of 16793 bp in length
* 10 61583 61682: gap of 100 bp
* 11 61683 80538: contig of 18856 bp in length
* 12 80539 80638: gap of 100 bp
* 13 80639 88966: contig of 8328 bp in length.
FEATURES
source
1..88966
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4q22-q24"
/clone="2185P9"
BASE COUNT 27101 a 15563 c 16501 g 29198 t 603 others
ORIGIN
Query Match 4.0%; Score 43.2; DB 2; Length 88966;
Best Local Similarity 52.2%; Pred. No. 5.9;
Matches 96; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```

```

QY 618 AGCAAAATGCTTTCTAATTCGAAGAAATTTAATGACTGCTTGATGGTCTAAGAAATCAG 677
Db 63048 AGAATATGAAGATTAGAAATTTAAAGCAGTGAAATAGAAATTTTAAAGAAATCAA 62989
QY 678 AAAGGAGACCGAAAGCTTCTAAGTAATCAAGAAAAAATCTTGACAATAGCTTTTCCATA 737
Db 62988 CCAAGCTAAGGTAGTGTTTTGAAGAGATCAACAACACTAACAGACATTTATCTATATG 62929
QY 738 GAGAAACCGCTGAGGAAGGAGTGCATACATAAATCCGATTAAGAAAGGAGGCTGCTA 797
Db 62928 AAGAAAAAAGAGAGAAAGATTCAATGGTATAAATACATATGAAAGGGGACATTA 62869
QY 798 CAAC 801
Db 62868 CAAC 62865

RESULT 23
AC093810
LOCUS AC093810/c 146671 bp DNA linear PRI 01-MAR-2002
DEFINITION Homo sapiens BAC clone RP11-354O24 from 4, complete sequence.
ACCESSION AC093810 AC023636
VERSION AC093810.4 GI:18042417
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 95063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 146671)
AUTHORS Kalicki,J. and Haglund,K.
TITLE The sequence of Homo sapiens BAC clone RP11-354O24
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 146671)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 146671)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 146671)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA.
On Jan 3, 2002 this sequence version replaced gi:16077045.
COMMENT ----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0354024
Drafting Center: WIBR
-----
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.
This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate

```

chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPL11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenos, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org> or VECTOR: pbace3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is AC087607; the clone sequenced to the right is RPL11-484F3, 2000 bp overlap. Actual start of this clone is at base position 1 of RPL11-354024; actual end is at base position 34976 of RPL11-484F3.

There is an ambiguous at base position 5636.

The sequence of AC023636 has been incorporated into AC093810.

#### FEATURES

source	Location/Qualifiers
repeat_region	1..146671
repeat_region	/organism="Homo sapiens"
repeat_region	/db_xref="taxon:9606"
repeat_region	/chromosome="4"
repeat_region	/map="4"
repeat_region	/clone="RPL11-354024"
repeat_region	/clone_lib="RPL11"
repeat_region	1..239
repeat_region	/rpt_family="ERV1"
repeat_region	341..503
repeat_region	/rpt_family="ERV1"
repeat_region	515..667
repeat_region	/rpt_family="ERV1"
repeat_region	681..1056
repeat_region	/rpt_family="ERV1"
repeat_region	1058..1534
repeat_region	/rpt_family="L1"
repeat_region	1644..1728
repeat_region	/rpt_family="MIR"
repeat_region	1746..2036
repeat_region	/note="match to EST T41120 (NID:9648688) ya31a05.s2"
repeat_region	1943..2237
repeat_region	/note="match to EST T40287 (NID:9647923) ya31e05.r2"
repeat_region	2872..2916
repeat_region	/rpt_family="AT_rich"
repeat_region	3714..4105
repeat_region	/rpt_family="L1"
repeat_region	4121..4623
repeat_region	/rpt_family="L1"
repeat_region	4199..4220
repeat_region	/rpt_family="AT_rich"
repeat_region	4638..4865
repeat_region	/rpt_family="L1"
repeat_region	4891..6849
repeat_region	/rpt_family="L1"
repeat_region	5591..5702
repeat_region	/rpt_family="(TA)n"
repeat_region	6532..6553
repeat_region	/rpt_family="AT_rich"
repeat_region	6748..6776
repeat_region	/rpt_family="(TAAA)n"
repeat_region	7416..7668
repeat_region	/rpt_family="L2"
repeat_region	8370..8483
repeat_region	/rpt_family="ACHobo"
repeat_region	8562..8954
repeat_region	/rpt_family="L1"
repeat_region	8968..9041
repeat_region	/rpt_family="MIR"
repeat_region	15112..15282
repeat_region	/rpt_family="MIR"
repeat_region	15296..15416
repeat_region	/rpt_family="L1"
repeat_region	15469..15706
repeat_region	/rpt_family="L1"
repeat_region	15759..15889
repeat_region	/rpt_family="Alu"
repeat_region	16136..16229
repeat_region	/rpt_family="L1"
repeat_region	17713..18027
repeat_region	/rpt_family="L1"
repeat_region	18028..18616
repeat_region	/rpt_family="L1"
repeat_region	18782..18840
repeat_region	/rpt_family="AT_rich"
repeat_region	19690..19949
repeat_region	/rpt_family="Alu"
repeat_region	20271..20872
repeat_region	/rpt_family="L2"
repeat_region	20903..21256
repeat_region	/rpt_family="L2"
repeat_region	21531..22139
repeat_region	/rpt_family="ERV1"
repeat_region	22360..22425
repeat_region	/rpt_family="L2"
repeat_region	22477..22646
repeat_region	/rpt_family="CRI"
repeat_region	22953..22993
repeat_region	/rpt_family="(TC)n"
repeat_region	24268..24346
repeat_region	/rpt_family="MIR"
repeat_region	24359..24406
repeat_region	/rpt_family="AT_rich"
repeat_region	26984..27033
repeat_region	/rpt_family="AT_rich"
repeat_region	27186..27799
repeat_region	/rpt_family="L2"
repeat_region	28814..28850
repeat_region	/rpt_family="AT_rich"
repeat_region	28906..29115
repeat_region	/rpt_family="L1"
repeat_region	29116..29149
repeat_region	/rpt_family="MaLR"
repeat_region	29150..29507
repeat_region	/rpt_family="L1"
repeat_region	29505..29527
repeat_region	/rpt_family="AT_rich"
repeat_region	30463..30505
repeat_region	/rpt_family="AT_rich"
repeat_region	30555..31651
repeat_region	/rpt_family="L1"
repeat_region	31663..32733
repeat_region	/rpt_family="L1"
repeat_region	32109..32133
repeat_region	/rpt_family="AT_rich"
repeat_region	32734..33108
repeat_region	/rpt_family="MaLR"
repeat_region	33109..33153
repeat_region	/rpt_family="L1"
repeat_region	33622..33908
repeat_region	/rpt_family="Alu"
repeat_region	33907..33936
repeat_region	/rpt_family="(TAA)n"
repeat_region	33937..33971

```
Query Match          4.0%; Score 43.2; DB 9; Length 146671;
Best Local Similarity 52.2%; Pred. No. 5.6;
Matches 96; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 618 AGGAAAAATGTTCTTAATTCAGAAATTTAATGACTGCTTGTGCTCTAAGAATACAG 677
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 42708 AGAATAATGAAGATTAGAATTAAAAGCAGTGAATATGAAATTTTAAAGAAATCAA 42649
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 678 AAAGGAGACCGGAAGCTTCTTAAGTAAATCAAGAAAAATCTTGGCAATAGCTTTCCCAT 737
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 42648 CCAAGCTAAAGGTAGGTTTTTTGAAAAGACTCAACAAAACTAACAGACATTTATCTATATG 42589
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 738 GAGAAACCGTGGAGGAAAGGAGTGCATACCAATAAATTCGATTAAAAAGGGAGGCTCTTA 797
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 42588 AAGAAAAAAGAAAGAAAGATTCAATGGATAAAATCACATATGAAAGGGGACATTA 42529
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 798 CAAC 801
      ||| |||
DB 42528 CAAC 42525

RESULT 24
AC107307/c
LOCUS      AC107307      158066 bp      DNA      linear      PRI 31-JUL-2002
DEFINITION Homo sapiens 3 BAC RP11-657G2 (Roswell Park Cancer Institute Human
BAC Library) complete sequence.
ACCESSION AC107307
VERSION    HTG.
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 158066)
            Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Burch,P., Burrett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
            Hollins,B., Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,O., King,L., Korvab,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Louisleg,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
            Neilson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N.,
            Nickerson,E., Nwokkwo,S., Ogih,M., Okwuonu,G., Oragunye,N.,
            Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
            Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
            Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
            Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E.,
            Sonalike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
            Tabar,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
            Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
            Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
            Warren,R., Washington,C., Watlington,S., Williams,G.,
```

Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,  
Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and  
Gibbs,R.

Direct Submission  
Unpublished  
2 (bases 1 to 158066)  
Worley,K.C.

Direct Submission  
Submitted (18-JAN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 158066)  
Worley,K.C.

Direct Submission  
Submitted (22-FEB-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 158066)  
Worley,K.C.

Direct Submission  
Submitted (28-MAR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

5 (bases 1 to 158066)  
Worley,K.C.

Direct Submission  
Submitted (02-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

6 (bases 1 to 158066)  
Worley,K.C.

Direct Submission  
Submitted (31-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Mar 28, 2002 this sequence version replaced gi:18449849.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
[gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.  
Overlapping clones are noted at the beginning and end of the  
Features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTS, GDB, and  
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2  
clones and 3 reads with no ambiguities. If the sequence quality for  
a region does not meet this standard, it will be indicated in the  
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated error rate less than 1 per 10,000 bases.  
Reports of lowest quality individual bases and measures of base  
quality are listed below. Description of the metrics can be found  
at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

FEATURES		QUALSTAT-REPORT.	
Source	Location/Qualifiers		
misc_feature	1. .158066	/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/chromosome="3"	
		/clone="RP11-657G2"	
	1. .2000	/note="overlaps bases 17989. .19988 of clone AC117506"	
repeat_region	136. .686	/function="clone overlap"	
		/rpt_family="L2"	
	717. .923	/rpt_family="L2"	
		complement(958. .1128)	
		/rpt_family="MER53"	
repeat_region	1252. .1799	/rpt_family="L2"	
		/rpt_family="L2"	
	3330. .3367	/rpt_family="AT_rich"	
	3519. .3940	/rpt_family="MSTC"	
	4242. .4287	/rpt_family="AT_rich"	
repeat_region	4421. .4479	/rpt_family="HERVL"	
		/rpt_family="MLT2B"	
		complement(4514. .4790)	
		/rpt_family="AluSx"	
	4895. .4924	/rpt_family="AT_rich"	
repeat_region	4951. .4985	/rpt_family="L1"	
		/rpt_family="L1"	
	5020. .5041	/rpt_family="AT_rich"	
	5079. .5504	/rpt_family="L1M2"	
	5508. .6426	/rpt_family="L1M2"	
repeat_region	6374. .6818	/rpt_family="L1"	
	6813. .8139	/rpt_family="L1MA4"	
	8140. .8445	/rpt_family="AluY"	
	8446. .8679	/rpt_family="L1MA4"	
	8686. .8997	/rpt_family="AluY"	
repeat_region	928. .9955	/rpt_family="AT_rich"	
	10136. .10175	/rpt_family="AT_rich"	
		complement(10229. .10588)	
		/rpt_family="THEIC"	
		complement(10700. .10996)	
repeat_region	11042. .11079	/rpt_family="AluY"	
		/rpt_family="AT_rich"	
	14756. .14800	/rpt_family="AT_rich"	
	16709. .17324	/rpt_family="L1MC4"	
	19258. .19280	/rpt_family="AT_rich"	
repeat_region	19658. .19680	/rpt_family="AT_rich"	
		complement(19837. .20082)	
	20605. .22071	/rpt_family="AluSx"	
		/rpt_family="MER52A"	
		complement(23423. .23578)	

repeat_region	repeat_region	Query Match	3.9%	Score 42.8;	DB 9;	Length 158066;
repeat_region	repeat_region	Best Local Similarity	45.3%	Pred. NO. 6.8;		
		Matches 155;	Conservative	0;	Mismatches 187;	Indels 0;
					Gaps 0;	



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 186158)  
Heilig, R., Petit, J.L., Vico, V., Dasilva, C., Robert, C., Wincker, P., Brothier, P., Catolico, L., Barbe, V., Pelletier, E., Artiguenave, F., Levy, M., Eckenberg, R., Bruls, T., Gherardinis, V., Cruaud, C., Gvayap, G., Saurin, W. and Weissenbach, J.  
Sequencing of the human chromosome 14  
Unpublished  
2 (bases 1 to 186158)  
Genoscope.  
Direct Submission  
Submitted (26-APR-2001) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
- Web : www.genoscope.cns.fr  
On Feb 28, 2001 this sequence version replaced gi:11611112.  
----- Genome Center  
Center: Genoscope / Centre National de Sequencage  
Center code: GS  
Web site: http://www.genoscope.cns.fr/  
Contact: SeqRef@genoscope.cns.fr  
-----

The following BAC sequence is oriented from the T7 to the SP6 end.  
Upstream BAC (overlapping the T7 end) : R-66E7  
Downstream BAC (overlapping the SP6 end) : R-423117 (AC=AL513142)  
----- Summary Statistics  
Assembly program: Phrap; version 2.0  
Quality coverage: 7.54x in Q20 bases; sum-of-contigs  
-----

Overall quality chart :  
Range : bases  
0 :  
1 : 9  
10 : 19  
20 : 29  
30 : 39  
40 : 49  
50 : 59  
60 : 69  
70 : 79  
80 : 89  
90 : 99

Percentage of bases with a quality value >= 40 : 99 %.

FEATURES  
source  
1. 186158  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/clone="R-60347"  
/clone\_lib="RPC1-11"  
146268. .146422  
/note="matching EMBL:238793  
RhdB:RH53848  
dbSTS:STS32572  
Identified using the e-PCR software (G. Schuler)"  
146274. .146413  
/note="matching EMBL:F01603  
RhdB:RH9897  
dbSTS:STS25118  
Identified using the e-PCR software (G. Schuler)"  
183972. .184175  
/note="matching EMBL:TI6335  
RhdB:RH53640  
dbSTS:STS42879  
Identified using the e-PCR software (G. Schuler)"  
183974. .184130  
/note="matching EMBL:TI6335  
RhdB:RH65383  
dbSTS:STS46312  
Identified using the e-PCR software (G. Schuler)"  
54122 a 35054 c 36149 g 60833 t

## ORIGIN

Query Match 3.9%; Score 42.8; DB 9; Length 186158;  
Best Local Similarity 50.5%; Pred. No. 6, 7; Mismatches 0; Gaps 0;  
Matches 104; Conservative  
QY 678 RAAGGAGACCGAAGCTTCTAAGTAAATCAAGAAAAATCTTGGACAATAGCTTTCCATA 737  
Db 61470 AAAGACATATACAGATGACAAATAGACACATGAATAATGTTAAAAAATCATTTAGTTATT 61529  
QY 738 GAGAAAACCGTGAGGAAGAGTGCATACCATTAATTCGGATTAAAAAGGAGTGTCTTA 797  
Db 61530 AGGGAATGCAAGAGACATGATGAGATCAATTAAGTACCTGTCAGATGACTGATAAATAA 61589  
QY 798 CAACAGCCATCATCGGAGCAACCTTCCTTTGTTCTTTGTAATGACAGAGATGAA 857  
Db 61590 AAAATGCCAAGATGTGGAACAACACAGCTCTTTTCATGCTGCTTATGAGAATGTAATAAT 61649  
QY 858 CAAGAGTAAGTTTTTTTGAAAAATTTCT 883  
Db 61650 GGTACAGCCATTCTAGGAACAGTTT 61675

## RESULT 27

AC112384 188385 bp DNA linear HTG 13-JUL-2002  
LOCUS Rattus norvegicus clone CH230-30L4, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
65 unordered pieces.  
DEFINITION AC112384  
ACCESSION AC112384  
VERSION AC112384.3 GI:21738458  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus

## REFERENCE

1 (bases 1 to 188385)  
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, P.R., Allen, C., Alsbrooks, S.L., Amaralunge, H.C., Are, J.R., Ayele, M., Banks, T., Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Eibaj, C., Escotto, M., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Foster, P., Frantz, P., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Gill, R., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homci, F., Howard, S., Huber, J., Huij, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, W., Liu, W., Loulseged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S., Oguh, M., Okwuonu, G., Oraguene, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojloban, I., Rolfe, M., Ruiz, S., Savery, G., Scher, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaike, I., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,



```

* 111170 111269: gap of unknown length
* 111270 113821: contig of 2552 bp in length
* 113821 113921: gap of unknown length
* 113921 117375: contig of 3454 bp in length
* 117375 117475: gap of unknown length
* 117475 121282: contig of 3807 bp in length
* 121282 121282: contig of 3807 bp in length

Query Match      3.9%; Score 42.8; DB 2; Length 188385;
Best Local Similarity 49.5%; Pred. No. 6.7;
Matches 110; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 668 AAGAAATCAGAGAGAGACCCGAAAGCTTCTTAAGTAATCAAGAAAAATCTTGACAAATAG 727
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171655 AAAAGAACTCAACCAAGATGAACAAAACAGGAGATAAACAAACAGAGAGATAAACT 171596

QY 728 GTTTCATAGACAAACCGTCGAGAAAGAGTGCATACCAATTCGATTAAAAAG 787
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171595 TTTATCCATACTAACAGAGTCCACAGAGTATCCAAATATCAATATCAAGAAAG 171536

QY 788 GGAGTGTTCACACGCCATATCAGTGGAGCAACCTCCCTTTGATCTCTTGATATGAC 847
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171535 GGAATATACACACAACTAAGAACTTCAAAAATATCAGATCCCTCTTCAAAATC 171476

QY 848 AGAAGATGACAAAGAGTAAAGTTTGTGAAATCTTAAATA 889
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171475 CTATGTTCAACAAACAGAAATTTGGATGAATGGACAAGA 171434

RESULT 28
AC092584/c
LOCUS      AC092584      159035 bp      DNA      linear      PRI 01-MAR-2002
DEFINITION Homo sapiens BAC clone RP11-20H7 from 2, complete sequence.
ACCESSION AC092584 AC015490
VERSION   AC092584.2 GI:16303525
KEYWORDS  HTG.
SOURCE    Homo sapiens.
  ORGANISM Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  AUTHORS Sulston, J.E. and Waterston, R.
  TITLE   Toward a complete human genome sequence
  JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE 95063792
  PUBMED  9847074
REFERENCE 2 (bases 1 to 159035)
  AUTHORS Armstrong, J., Kozlowski, A., Hawkins, M. and Spalding, L.
  TITLE   The sequence of Homo sapiens BAC clone RP11-20H7
  JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 159035)
  AUTHORS Waterston, R.H.
  TITLE   Direct Submission
  JOURNAL Submitted (19-JUL-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE 4 (bases 1 to 159035)
  AUTHORS Waterston, R.H.
  TITLE   Direct Submission
  JOURNAL Submitted (20-OCT-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE 5 (bases 1 to 159035)
  AUTHORS Waterston, R.H.
  TITLE   Direct Submission
  JOURNAL Submitted (23-OCT-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE 6 (bases 1 to 159035)
  AUTHORS Waterston, R.
  TITLE   Direct Submission
  JOURNAL Submitted (01-MAR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT   On Oct 20, 2001 this sequence version replaced gi:14916169.

```

----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@wustl.edu](mailto:sapiens@wustl.edu)

----- Summary Statistics  
Center project name: H\_NH0020H07  
Drafting Center: WIBR  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>  
VECTOR: pBACE3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-138C13, 2000 bp overlap; the clone sequenced to the right is RP11-21L7. Actual start of this clone is at base position 1 of RP11-20H7; actual end is at base position 159035 of RP11-20H7.

Data from AC068537, AC073994, and AC016827 was used to finish this clone, AC092584.

The sequence of AC015490 has been incorporated into AC092584.

#### FEATURES

source	Location/Qualifiers
1..159035	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-20H7"
	/clone_lib="RP11-11"
repeat_region	1..108
	/rpt_family="L1"
repeat_region	85..149
	/rpt_family="L1"
repeat_region	148..550
	/rpt_family="L1"
repeat_region	344..387
	/rpt_family="AT-rich"
repeat_region	550..2451
	/rpt_family="L1"
repeat_region	2031..2068
	/rpt_family="(TG)n"
repeat_region	3780..4131
	/rpt_family="MaLR"
repeat_region	4308..4401
	/rpt_family="L2"



```
repeat_region 5512..5561 /rpt_family="Mariner"
repeat_region 5545..5567 /rpt_family="AT_rich"
repeat_region 6515..7739 /rpt_family="L1"
repeat_region 7582..7622 /rpt_family="L1"
repeat_region 8664..8814 /rpt_family="TCTA)n"
repeat_region 9632..9817 /rpt_family="MIR"
repeat_region 9948..9990 /rpt_family="L1"
repeat_region 10104..10426 /rpt_family="(TCCA)n"
repeat_region 10466..10564 /rpt_family="L1"
repeat_region 10574..10605 /rpt_family="(TTTA)n"
repeat_region 10601..11026 /rpt_family="(TTTA)n"
repeat_region 11062..11958 /rpt_family="L1"
repeat_region 11334..11362 /rpt_family="L1"
repeat_region 11987..12289 /rpt_family="AT_rich"
repeat_region 12777..12812 /rpt_family="AT_rich"
repeat_region 12799..13336 /rpt_family="L1"
repeat_region 13270..13291 /rpt_family="AT_rich"
repeat_region 13365..13877 /rpt_family="L1"
repeat_region 13879..14061 /rpt_family="G-rich"
repeat_region 14072..14232 /rpt_family="L1"
repeat_region 14231..14412 /rpt_family="L1"
repeat_region 14391..14436 /rpt_family="AT_rich"
repeat_region 14492..15506 /rpt_family="L1"
misc_feature 15516..15895 /note="similar to Homo sapiens EST BE565142 (NID:59808862)"
repeat_region 16673..16839 /rpt_family="Alu"
repeat_region 17085..18013 /rpt_family="L1"
repeat_region 18027..18380 /rpt_family="L1"
repeat_region 18381..18406 /rpt_family="(T)n"
repeat_region 18684..18988 /rpt_family="L1"
repeat_region 19010..19060 /rpt_family="Mariner"
repeat_region 19069..19154 /rpt_family="L1"
repeat_region 19830..19852 /rpt_family="(GAA)n"
repeat_region 21727..21845 /rpt_family="T-rich"
repeat_region 22856..22888 /rpt_family="(TTTC)n"
repeat_region 26258..26691 /rpt_family="L1"
repeat_region 27951..28074 /rpt_family="L1"
```

```
repeat_region 28081..29567 /rpt_family="L1"
repeat_region 29568..29848 /rpt_family="Alu"
repeat_region 29849..32251 /rpt_family="L1"
repeat_region 31201..31310 /rpt_family="n-rich"
misc_feature 32226..33096 /note="similar to Sus scrofa EST B1184750 (NID:914659159)"
misc_feature 32227..32588

Query Match 3.9%; Score 42.4; DB 9; Length 159035;
Best Local Similarity 52.8%; Pred. No. 8.5; Mismatches 1; Gaps 1;
Matches 114; Conservative 0;

QY 672 AATCAGAAAGGAGCCGAAGCTTCTAAGTAATCAAGAAAAATCTGGCAATAGCTTT 731
DB 21746 AAAAATCAATGAACAAAGAGTGTGTTTTTAAATTAACAAATTCGCAA-ACTTTA 21688
QY 732 TCCATAGAGAAACCCGTGAGGAAGAGTCGATACCATAAATTCGGATTAAAAAGGGAG 791
DB 21687 GCCAGACAAAGAAAAAATGAGAAAGGTCACAAATAAATAACAGAAATGAAATPGGAG 21628
QY 792 TGCATTACAAACAGGCATATCAGTGGAGCAACCTTCCTTTGATTCTTTGATAATGACAGAA 851
DB 21627 AGGTTACAACCTGCTATCCACAGAAATTCAAAGGATTATTGTGGCTCTATGAGCAACGTA 21568
QY 852 GATGAACAAAGAGTAAGTTTTTGAAAAATTCFAAAA 887
DB 21567 AGTCACAAAGTGAAAAATCTAGAAGAAATGCTAAA 21532

RESULT 29
AF268060 2706 bp DNA linear BCT 24-MAY-2001
LOCUS
DEFINITION
Candidatus Carsonella ruddii natural-host Aphalaroida inermis RNA
polymerase beta subunit (rpoB) and RNA polymerase beta-prime
subunit (rpoC) genes, partial cds.
ACCESSION AF268060
VERSION AF268060.1 GI:14193385
KEYWORDS
SOURCE
ORGANISM
Candidatus Carsonella ruddii.
Bacteria; Proteobacteria; gamma subdivision; Candidatus Carsonella.
REFERENCE
1 (bases 1 to 2706)
AUTHORS
Thao,M.L., Clark,M.A., Burckhardt,D.H., Moran,N.A. and Baumann,P.
TITLE
Phylogenetic analysis of vertically transmitted psyllid
endosymbionts (Candidatus Carsonella ruddii) based on atpAGD and
rpoC; comparisons with 16S-23S rDNA-derived phylogeny
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 2706)
AUTHORS
Thao,M.L., Clark,M.A., Burckhardt,D.H., Moran,N.A. and Baumann,P.
TITLE
Direct Submission
JOURNAL
Submitted (16-MAY-2000) Microbiology Section, University of
California, Davis, One Shields Avenue, Davis, CA 95616-8665, USA
FEATURES
Location/Qualifiers
1..2706
/organism="Candidatus Carsonella ruddii"
/specific_host="Aphalaroida inermis"
/db_xref="taxon:114186"
<1..184
/ene="rpoB"
<1..184
/ene="rpoB"
/codon_start=2
/trans_start=11
/product="RNA polymerase beta subunit"
/protein_id="AAK55948.1"
/db_xref="GI:14193386"
/translation="YGAANILREMLTIKSDDLLEGRTCLFKNIINDIHYINFGIPESFO
ILIQFIKSLCPDIIV"
181..>2706
gene
```

CDS

/gene="rpoC"  
181..>2706  
/gene="rpoC"  
/codon\_start=1  
/transl\_table=11  
/product="RNA polymerase beta-prime subunit"  
/protein\_id="AAK5949.1"  
/db\_xref="GI:14193387"  
/translation="MKIKKIIILKASPNKIKNWSYGEIKSNFNFKNLKPKNGLF  
CKIFSYNICECKKCNKNNFLNLRKSHSYRGHKLYYPALHWFKNVNNI  
SLISLSEKILKIVLSEKPKKYSLSLKNNILNCYLVSAGAKLKM  
SNEILLDCVLLKLLNCSENLPLDYLDKTKIYFVLSGNPEWPCULIPVSP  
KTRPLIPNLNICKYALSDNLYKKVDNRILKKKILKPKQIINERILLOSINA  
LFDNKEIENPLTNSKRVLSFSNSIKGIRFQNLGKRVDFSGRSVLSYDPELAL  
YCKPILIGIEFKPIFNKLAKKIIENINFLDEFYLNKKKESINILADCKNKT  
LNLRAPTHMGISQKILTDQTKIIRHPLICLSADFDGQMAHLPIITINAQVE  
SNYLLSNLISPSNGEPIIIPQDVIWGIYCLTNVNNYKILPFSFDFVKNFKN  
NKAENVLNKNYFIKTVGRIFYYILKLTININFKVIRKKDLIYIKIYEFY  
GVYKTKILDNKKIGFFFTFISYDLPDKIENILKLLKTKNNKDNFINI  
NVVELLIENIIKKIKTNKKNNLFLMDSGRSLOIKQLLGRGPFSGNGBI  
IIDPLDNKGLSMKNYFISFGARKGLTDSLKTANSGLYLRKLVDMOGIVYKI  
DCGVYGLIEILKNNFYLLYKIFGRIDISIFPKNLIKLNKNTLNKRIIPLIK  
KKNKIYIIRSVLHCVSIRGICVLCYGLISNYKLINIGTPIGII"  
BASE COUNT 1199 a 188 c 266 g 1053 t  
ORIGIN

Query Match  
Best Local Similarity 3.9%; Score 42.2; DB 1; Length 2706;  
Matches 104; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 830 TGATTCCTTCATACAGAGAGTCAACAGAGTAAAGTCTTAAATA 889  
DB 2124 TAATTTATTAATGTTAGTTCGGTCAAGAGAGTATTTTACAATCAACAAAT 2183  
QY 890 GAATGATTCCTCGACCTAAAGCTGCTCTCTTTATTTCCITTTGAGATGAT 949  
DB 2184 ATTAGTTTATAGAGGATTTTTCACAAATCTAAGTGAATATTTATGATCCAAAT 2243  
QY 950 CGGTAGTAGAATATGGGTGAGATGACGCTGGAGTGAATGCGTTCGTGGAAGAAC 1009  
DB 2244 AGATATTTTAAATAGGATTAAGTATGAAATATTTTATTAAGCTCTTTGGAGCAAG 2303  
QY 1010 AGGAAAGATGATGCTCAGCTTTAGA 1036  
DB 2304 AAAAGGATTAAGTACTCTCTTTAAA 2330

RESULT 30  
AC1010342/c  
LOCUS  
DEFINITION Homo sapiens chromosome 5 clone CTD-2015B23, complete sequence.  
ACCESSION AC1010342  
VERSION AC1010342.5 GI:13876487  
KEYWORDS HTG  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
DOE Joint Genome Institute and Stanford Human Genome Center.  
Direct Submission  
TITLE  
2 (bases 1 to 126010)  
DOE Joint Genome Institute.  
REFERENCE  
AUTHORS  
JOURNAL  
TITLE  
2 (bases 1 to 126010)  
DOE Joint Submission  
Direct Submission  
Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 126010)  
DOE Joint Genome  
DOE Joint Submission  
Direct Submission  
Submitted (28-APR-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
4 (bases 1 to 126010)  
REFERENCE

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (01-MAY-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
COMMENT On Apr 28, 2001 this sequence version replaced gi:13699503.  
Draft Sequence produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >40 99.9% of Sequence;  
Estimated Total Number of Errors is 0.1.  
STS Content  
SHGC-104285\_G58244  
WI-4392\_G02894.  
FEATURES  
source  
1. 126010  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTD-2015B23"  
BASE COUNT 41323 a 21680 c 22194 g 40813 t  
ORIGIN

Query Match  
Best Local Similarity 3.9%; Score 42.2; DB 9; Length 126010;  
Matches 188; Conservative 0; Mismatches 243; Indels 0; Gaps 0;

QY 465 AAGGACCTGCGCCCTAAAGGTCATGGATGAGATCAGGAGTTAAGGAGAAATATCTTTAT 524  
DB 56662 AGAATTTTGGCATTAACATAAACATGTTTGCATGATGATATGTTGCACCTGCCCTCA 56603  
QY 525 CGAGAGTTCGCATTACAGAGCGCAATGGGAAAGTAAAGCTGCTGCACAGAGACAAGTATT 584  
DB 56602 AAAGATTAACAGGAAATAGCAATTTAAATATACGGACATCTACAGAGCAACAACT 56543  
QY 585 GATGGGAGCTACAAATCCCTGAAACGGGATAATAGAAATAATGCTTTCTTAATTCGAAGA 644  
DB 56542 TAATAGATAACAAATTTACTCAAAATTTTAAATATCAATATATATATATATTTGTTAAGA 56483  
QY 645 ATTTAATGACCTGCTTGATGCTTAAGAAATACAGAAAGGAGACCGGAAAGCTTCTTAAGTAAT 704  
DB 56482 AAGTCAATTTGATTAGAAATTTGAAAGATAGTGTATATATATACCAATAAATTTGGAAT 56423  
QY 705 CAAGAAAAATCTTTGGCAATAGCTTTTCCATAGAGAAAAACCGTGAGGAAAGAGTGCAT 764  
DB 56422 TTATATAAAGCCCTACTATAAATCTTTGTTTAAAGAAATATGTTTACAAAGGAAATATGG 56363  
QY 765 ACCATAAATTCGGATTTAAAGGAGTGTCTTACAAACAGCCATATATCAGTGAGCAACCTT 824  
DB 56362 CCTCAACTGAAGGACATTAAGTATGCTATATAAACCATGAGATACCTGAGTATAGTGA 56303  
QY 825 CCTTTTGATCTTTTGATAATGACAGAGATGAACAAAGAGTAAGTTTTTGAATAATTTCTA 884  
DB 56302 CACATGGAGGTATTATTACAGCCTTAACCCATCGCTTCAGAAAACGTTTGAANAATCACTA 56243  
QY 885 AAATAGAAATG 895  
DB 56242 AGTTTGAATG 56232

RESULT 31  
AC108007/c  
LOCUS  
DEFINITION Homo sapiens chromosome 11 clone CTD-3038L12 map 11, \*\*\* SEQUENCING  
IN PROGRESS \*\*\*, 8 ordered pieces.  
ACCESSION AC108007  
VERSION AC108007.4 GI:22474954  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 141635)

AUTHORS Birren,B., Nusbaum,C. and Lander,E.  
 TITLE Homo sapiens chromosome 11, clone CTD-3038L12  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 141635)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferrel,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kellis,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Stojanovic,N., Severi,P., Spencer,B., Stange-Thomann,N., Theodores,J., Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodores,J., Topham,K., Travers,M., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
 Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 141635)  
 Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kellis,C., Landers,T., Levine,R., Lindblad-Rohat,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,C., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severi,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Testaye,S., Theodores,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Willson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission  
 Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Aug 24, 2002 this sequence version replaced gi:20336822.  
 All repeats were identified using RepeatMasker:  
 Smit,A.F.A. & Green,P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Genome Institute  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 Project Information  
 Project name: L24596  
 Project clone name: 3038\_L12

\*\*\*\*\*  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 8 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* been provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

1 92004: contig of 92004 bp in length  
 \* 92005 92104: gap of 100 bp  
 \* 92105 117227: contig of 25123 bp in length  
 \* 117228 117327: gap of 100 bp  
 \* 117328 120434: contig of 3107 bp in length  
 \* 120435 120534: gap of 100 bp  
 \* 120535 126202: contig of 5668 bp in length  
 \* 126203 126302: gap of 100 bp  
 \* 126303 129158: contig of 2856 bp in length  
 \* 129159 129258: gap of 100 bp  
 \* 129259 131561: contig of 2703 bp in length  
 \* 131562 132061: gap of 100 bp  
 \* 132062 135739: contig of 3678 bp in length  
 \* 135740 135836: gap of 100 bp  
 \* 135840 141635: contig of 5796 bp in length.

Location/Qualifiers  
 1. 141635  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="11"  
 /map="11"  
 /clone="CTD-3038L12"  
 /clone\_lib="CIRD2 Human BAC"  
 /base\_count 42227 a 27391 c 27480 g 43740 t 797 others  
 ORIGIN

Query Match 3.9%; Score 42.2; DB 2; Length 141635;  
 Best Local Similarity 43.6%; Pred. No. 9.6;  
 Matches 188; Conservative

QY 465 AAGGCACGTCGCCCTAAAGGTCATGATGAGATCAGATCAGGAGTTAAGGAGAAATATCTTTAT 524  
 Db 119252 AAGAAATTTGCATAAATAAACAATGTTTGCATGATGAATATCTTGCACCTCTGCCCTCA 119193  
 QY 525 CGAGAGTTTCGCATTTACAGAGGCAATGGGAAAGTAAAGCTCTGCAACAGAGAGCAAGTAAT 584  
 Db 119192 AAGAGATACAGGAAATAGCAATTTAAATAATATACGCACATCTACAGAGCAACAACCT 119133  
 QY 585 GATGGGAGCTACAAATCTTGAACGGGATATAGAAAATGTTTCTTAATTCGAAGA 644  
 Db 119132 TAATAGATACAAATTTACTCAAAATTTAAATATCAATATATATATATATATATATAT 119073  
 QY 645 ATTTAATGACTGCTTGATGCTCTAAGAAATCAGAAAGGAGACCGGAAGCTTCTAAGTAAT 704  
 Db 119072 AAGTCAATTTAGTAATTTAGAAATTTAGATAGTAGTGTATATATACCAATATATTTGGAAT 119013  
 QY 705 CAAGAAAAATCTTGGACATAGCTTTTCCATAGAGAAACCGTGAGGAAGAGGAGTCGAT 764  
 Db 119012 TTAATAAAGCCCTACTAAATAACTTTTGGTTTAAAGAAATATGTTTCAAGGAATATGG 118953  
 QY 765 ACCATAAATTCGATTTAAAGAGGAGTGCTTACACAGCCATATCAGTGGAGCAACCTT 824  
 Db 118952 CCTCACTCAAGGACATTAAGTATGCTTATATAAAACCATGAGATACGTGAGTATAGTGA 118893  
 QY 825 CCCTTTGATTTCTTTGATTAATGACAAAGAGTGAACAAGAGTAGTAAAGTTTGTGAAAAATTTCTA 884  
 Db 118892 CACATGAGGTATATTATTACAGCCCTAAACCCATCGCTTTCAGAAAACGTTTGAAAAATCACTA 118833

QY 885 AAATAGAAATG 895  
 Db 118832 AGTTTGAATG 118822

RESULT 32  
 AL807375/c 192168 bp DNA linear HTG 17-JUL-2002  
 LOCUS Mus musculus chromosome X clone RP23-247H23, \*\*\* SEQUENCING IN  
 DEFINITION PROGRESS \*\*\*, 8 unordered pieces.  
 AL807375  
 ACCESSION AL807375.5 GI:21738714  
 VERSION HTG: HTGS\_PHASE1.  
 KEYWORDS Mus musculus.  
 SOURCE Mus musculus  
 ORGANISM

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Plumb,B.
Direct Submission
Submitted (14-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jul 12, 2002 this sequence version replaced gi:21727570.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BM247H23
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 189470 bases at least Q40
Consensus quality: 190260 bases at least Q30
Consensus quality: 190881 bases at least Q20
Insert size: 191468; sum-of-contigs
Insert size: 192804; 3.1% error; agarose-fp
Quality coverage: 6.94x in Q20 bases; sum-of-contigs Quality
coverage: 6.89x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 55213: contig of 55213 bp in length
* 55214 55313: gap of 100 bp
* 55314 66635: contig of 11322 bp in length
* 66636 66735: gap of 100 bp
* 66736 73081: contig of 6346 bp in length
* 73082 73181: gap of 100 bp
* 73182 118845: contig of 45664 bp in length
* 118846 118945: gap of 100 bp
* 118946 130156: contig of 11211 bp in length
* 130157 130256: gap of 100 bp
* 130257 135973: contig of 5717 bp in length
* 135974 136073: gap of 100 bp
* 136074 138537: contig of 2864 bp in length
* 138538 139037: gap of 100 bp
* 139038 192168: contig of 53131 bp in length.
*
* Location/Qualifiers
* 1..192168
* /organism="Mus musculus"
* /db_xref="taxon:10090"
* /chromosome="X"
* /clone="RP23-247H23"
* /clone_lib="RPCI-23"
* 1..55213
* /note="assembly_fragment:00451
* fragment_chain:1
* clone_end:SP6
* vector_side:left"
* 55314..66635
* /note="assembly_fragment:02335
* fragment_chain:1"
* 66736..73081
* /note="assembly_fragment:00648
* fragment_chain:1"
* 73182..118845
* /note="assembly_fragment:01518
* fragment_chain:1"
* 118946..130156
* /note="assembly_fragment:02035

```

```

fragment_chain:1"
130257..135973
/note="assembly_fragment:01769
fragment_chain:1"
136074..138937
/note="assembly_fragment:01554
fragment_chain:1"
139038..192168
/note="assembly_fragment:02173
fragment_chain:1"
clone_end:r7
vector_side:right"
59494 t 703 others
ORIGIN
Query Match 3.9%; Score 42.2; DB 2; Length 192168;
Best Local Similarity 55.0%; Pred. No. 9.3;
Matches 83; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 188 AATCTTAAGGAGGACACAGCTAGCTTATCAATGAAGCGTGTCCTGAGGAGAA 247
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155143 AATGTTGAAGCAGCAAAAATCTCAATAACATTTAAAGCAAGGTCGAGAGAA 155084
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 248 CTGATAAGACGGCTCCATATGCTTTCTTCCTTCATGTTTTCATTAAGTCGCTTCA 307
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155083 CACATAGTACTTCAGGAATCAATTTTTCATGGAATAGGGGATATCCTTAA 155024
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 308 ACAGTAACCAAAAGCAGCGTACATAAGCGA 338
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155023 AATGAACCTTAGCTGCGCTTCATTAGAGA 154993
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 33
AL831718
LOCUS
DEFINITION
AL831718 Mus musculus chromosome X clone RP23-146020, *** SEQUENCING IN
PROGRESS ***, 5 unordered pieces.
AL831718
AL831718.5 GI:22213735
VERSION
HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
house mouse.
SOURCE
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Clark,S.
Direct Submission
Submitted (08-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Aug 13, 2002 this sequence version replaced gi:21953128.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BMI46020
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 202322 bases at least Q40
Consensus quality: 202353 bases at least Q30
Consensus quality: 202377 bases at least Q20
Insert size: 202401; sum-of-contigs
Insert size: 197576; 6.0% error; agarose-fp
Quality coverage: 13.08x in Q20 bases; sum-of-contigs Quality
coverage: 13.69x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

```

\* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 92824: contig of 92824 bp in length  
 92825 92924: gap of 100 bp  
 92925 102534: contig of 9610 bp in length  
 102535 102634: gap of 100 bp  
 102635 109946: contig of 7312 bp in length  
 109947 110046: gap of 100 bp  
 110047 112095: contig of 2049 bp in length  
 112096 112195: gap of 100 bp  
 112196 202801: contig of 90606 bp in length.

FEATURES  
 Source  
 1. .202801  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="X"  
 /clone="RP23-146020"  
 /clone\_lib="RPCT-23"  
 1. 92824  
 /note="assembly\_fragment:01268"  
 /fragment\_chain:1  
 92825..102534  
 /note="assembly\_fragment:01238"  
 /fragment\_chain:1  
 102635..109946  
 /note="assembly\_fragment:03032"  
 /fragment\_chain:1  
 110047..112095  
 /note="assembly\_fragment:01164"  
 /fragment\_chain:1  
 112196..202801  
 /note="assembly\_fragment:00215"  
 /fragment\_chain:1  
 64339 a 38445 c 38425 g 61192 t 400 others

Query Match 3.9%; Score 42.2; DB 2; Length 202801;  
 Best Local Similarity 55.0%; Pred. No. 9.2;  
 Matches 83; Conservative 0; Mismatches 68; Indels 0; Gaps 0;  
 QY 188 AATTCCTAAAGGAGAACACGTACCTTATCAATGAAAGCGTGCCCAAGTTCCTGAGGAGAA 247  
 Db 190083 AATGTTGAAGCAGAAAAATCTCAATAACATTAAGCAAGGTCAAGTCTCGAGAAGAA 190142  
 QY 248 CTGATAGAACGGCTCCATATGCTTTTCTTCCTTCATGCTTTTCAATAAGTCCGCTTTCA 307  
 Db 190143 CACATAGTTACTTTCAGGAATCAATTTTTTTTTCATGGAATAAGGCGATATGCTTTAA 190202  
 QY 308 ACAGTAACCAAGACGCTACATAGCGA 338  
 Db 190203 AATGAACCTTACCTGCTTCATTAGAGA 190233

RESULT 34  
 AP000988/c 266050 bp DNA linear BCT 11-OCT-2001  
 LOCUS Sulfolobus tokodaii genomic DNA, complete sequence, section:8/10.  
 DEFINITION AP000988 BA000023  
 ACCESSION AP000988.1 GI:15622956  
 VERSION  
 KEYWORDS  
 SOURCE  
 Sulfolobus tokodaii (strain:7) DNA.

ORGANISM  
 Sulfolobus tokodaii  
 Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 Sulfolobus.

REFERENCE  
 AUTHORS  
 1 Kwarabayasi,Y., Hino,Y., Horikawa,H., Jin-no,K., Takahashi,M.,  
 Sekine,M., Baba,S., Anai,A., Kosugi,H., Hosoyama,A., Fukui,S.,  
 Nagai,Y., Nishijima,K., Otsuka,R., Nakazawa,H., Takamiya,M.,  
 Kato,Y., Yoshizawa,T., Tanaka,T., Kudoh,Y., Yamazaki,J.,  
 Kushida,N., Oguchi,A., Aoki,K., Masuda,S., Yanagii,M.,  
 Nishimura,M., Yamagishi,A., Oshima,T. and Kikuchi,H.

TITLE  
 JOURNAL  
 MEDLINE  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 Complete genome sequence of an aerobic thermoacidophilic  
 Crenarchaeon, Sulfolobus tokodaii strain7  
 DNA Res. 8 (4), 123-140 (2001)  
 21456156  
 2 (bases 1 to 266050)  
 Kwarabayasi,Y., Tanaka,T., Hino,Y. and Kikuchi,H.  
 Direct Submission  
 Submitted (22-DEC-1999) Yutaka Kwarabayasi, National Institute of  
 Technology and Evaluation, Biotechnology Center; Nishihara 2-49-10,  
 Shibuya-ku, Tokyo 151-0066, Japan (E-mail:kyutaka@nite.go.jp,  
 URL:http://www.bio.nite.go.jp/, Tel:81-3-3481-8951,  
 Fax:81-3-3481-8424)  
 Kwarabayasi,Y. is officially affiliated with the National  
 Institute of Advanced Industrial Science and Technology, Tsukuba,  
 Ibaraki 305-0046, Japan  
 Yamagishi,A. and Oshima,T. are at Tokyo University of Pharmacy  
 and Life Science, Hachioji, Tokyo 192-0392, Japan  
 The other authors are at the National Institute of Technology and  
 Evaluation, Nishihara 2-49-10, Shibuya-ku, Tokyo 151-0066, Japan  
 Homology analysis is performed by Smith-Waterman algorithm against  
 Genbank release 109; EMBL release 36.0; SwissProt release 38.0;  
 PIR-Protein release 62.0; and OWL release 31.4.  
 E-mail address for comments and questions: kyutaka@nite.go.jp  
 ORF organization, sequence alignment and more information are  
 available at W.W.W. site of Biotechnology Center,  
 URL: http://www.bio.nite.go.jp/dogan/genome\_list-e.html/.

FEATURES  
 source  
 1. .266050  
 /organism="Sulfolobus tokodaii"  
 /strain="7"  
 /db\_xref="taxon:111955"  
 288. .830  
 /gene="ST1855"  
 288. .830  
 /gene="ST1855"  
 /note="OWL:AAD0XP24H1 percent identity: 83.432 in 169aa."  
 /codon\_start=1  
 /transl\_table=11  
 /product="180aa long hypothetical protein"  
 /protein\_id="BAB66948.1"  
 /db\_xref="GI:15622957"  
 /translation="MSKINFEKNATRELYIPIREAVGMFLDGLGRKAVLNPAKILD  
 PNSSVFGKLYNLEPHAGPKPFLIATLENHAFDVFTELTIFSIEIAGLFLVIGLL  
 TRLAAGAAVMAIGHAPAYMGLSGCEDEWQIGSLITAGAIIVMLTGAAGVWGLDYFLY  
 KKFQDRIANVPILKWIKLN"  
 824. .1342  
 /gene="ST1856"  
 824. .1342  
 /gene="ST1856"  
 /note="OWL:AAD0XP24H2 percent identity: 68.590 in 156aa."  
 /codon\_start=1  
 /transl\_table=11  
 /product="172aa long hypothetical protein"  
 /protein\_id="BAB66947.1"  
 /db\_xref="GI:15622958"  
 /translation="MVRKMKVAIVGIIFALFTVGWLTGQWYNNVGVLPNNSKL  
 PKLNTVITAYNTPQGLCLIMNTIDGDAYPASAPIMEISNGTWHEIENSSQIAND  
 TVKLIQAPWNANKSVNNYSGFVILGSEGFHLIKLGLHLSPGKRYEKLTYPAISS  
 NRQAIAYFTISS"  
 1378. .2772  
 /gene="ST1857"  
 1378. .2772  
 /gene="ST1857"  
 /note="PIR:D64486 percent identity: 34.286 in 105aa."  
 /codon\_start=1  
 /transl\_table=11  
 /product="464aa long conserved hypothetical protein"  
 /protein\_id="BAB66948.1"  
 /db\_xref="GI:15622959"  
 /translation="MENYSKEIFERASOTYSDGILGLIKRGNKILGIVKDDIYRVE  
 IDLSKSGKCCCLGNCPLHAIKMSYKPGYIDFDSLENKITEINKRELKGLIVTL  
 IEKFPFIANYIPIENAKISLSEYINLIKNGPENIVNSFTDFLINNRKINKDDIFI  
 ILDTIASCKSKCFYNFIERDENLMTLANILLEKEVEDDKIKLEIKDKYGN



AUTHORS  
TITLE Toward a complete human genome sequence  
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
MEDLINE 99063792  
PUBMED 9847074

REFERENCE  
2 (bases 1 to 98697)  
Ozanich,A. and Mead,K.  
The sequence of Homo sapiens PAC clone RP4-673M15  
Unpublished (2001).

AUTHORS  
JOURNAL  
REFERENCE  
3 (bases 1 to 98697)  
Waterston,R.H.  
Direct Submission  
Submitted (12-JUN-1998) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

AUTHORS  
JOURNAL  
REFERENCE  
4 (bases 1 to 98697)  
Waterston,R.H.  
Direct Submission  
Submitted (14-MAY-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

AUTHORS  
JOURNAL  
REFERENCE  
5 (bases 1 to 98697)  
Waterston,R.  
Direct Submission  
Submitted (22-OCT-1999) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

AUTHORS  
JOURNAL  
REFERENCE  
6 (bases 1 to 98697)  
Waterston,R.  
Direct Submission  
Submitted (21-DEC-1999) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

AUTHORS  
JOURNAL  
REFERENCE  
7 (bases 1 to 98697)  
Waterston,R.  
Direct Submission  
Submitted (04-JUN-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Jun 4, 2002 this sequence version replaced gi:4827328.

----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@wustl.edu](mailto:sapiens@wustl.edu)  
----- Summary Statistics  
-----  
Center project name: H\_DJ0673M15  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see [http://www.nhgri.nih.gov/DIR/GB/CHR7\\_send\\_mailto:greennhgri.nih.gov](http://www.nhgri.nih.gov/DIR/GB/CHR7_send_mailto:greennhgri.nih.gov), or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
This clone was derived from human PAC library RPCI-4, prepared by Pieter de Jong and coworkers at <http://www.chori.org> using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994).

The library is from one male donor.  
The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.  
VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the right is DJ0647J21. Actual start of this clone is at base position 1 of DJ0673M15.

Base positions 19689 to 19719 of the submitted sequence is represented by single chemistry only.

FEATURES	source
1..98697	Location/Qualifiers
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="7"	
/map="7p13-p11.2"	
/clone="RP4-673M15"	
/clone.lib="RPCI-4"	
4..568	
/note="similar to EST AW131540 (NID:g6133147) xf31b08.x1"	
4..568	
/note="similar to EST BF026668 (NID:g10734380)"	
4..548	
/note="similar to EST AA775847 (NID:g2835181) adi7e01.s1"	
4..540	
/note="similar to EST BE727822 (NID:g10141814)"	
4..318	
/note="similar to EST AI017276 (NID:g3231612) ov27f07.x1"	
4..310	
/note="similar to EST AA216094 (NID:g1816041)"	
33..552	
/note="similar to EST BE075304 (NID:g8424032)"	
177..568	
/note="similar to EST BE588888 (NID:g9841927)"	
216..568	
/note="similar to EST BE256638 (NID:g9127101)"	
254..568	
/note="similar to EST AA057332 (NID:g1550170) zf56all.r1"	
554..857	
/rpt_family="Alu"	
862..1003	
/note="similar to EST BE727822 (NID:g10141814)"	
886..1080	
/note="similar to EST AA057332 (NID:g1550170) zf56all.r1"	
886..1080	
/note="similar to EST BF026668 (NID:g10734380)"	
886..1079	
/note="similar to EST BE588888 (NID:g9841927)"	
886..1075	
/note="similar to EST BE256638 (NID:g9127101)"	
886..922	
/note="similar to EST AW131540 (NID:g6133147) xf31b08.x1"	
1381..1662	
/rpt_family="L1"	
1663..1954	
/rpt_family="Alu"	
1673..1727	
/note="similar to EST A1147019 (NID:g3674701) qa82a12.x1"	
1959..2254	
/rpt_family="Alu"	
2255..2303	
/rpt_family="L1"	
2329..2467	
/rpt_family="L1"	
2468..2522	
/rpt_family="L1"	
2912..3207	
/rpt_family="Alu"	
3239..3404	
/rpt_family="Alu"	
3625..3935	





```
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 59885 bases at least Q40
Consensus quality: 63135 bases at least Q30
Consensus quality: 65093 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
      (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 69 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1125: contig of 1125 bp in length
1126: gap of unknown length
1246: contig of 1020 bp in length
2345: gap of unknown length
3896: contig of 1551 bp in length
3897: gap of unknown length
3997: contig of 1576 bp in length
5572: gap of unknown length
5673: contig of 1026 bp in length
6698: gap of unknown length
6799: contig of 1343 bp in length
8141: gap of unknown length
8242: contig of 1269 bp in length
9511: gap of unknown length
9610: contig of 1495 bp in length
11105: gap of unknown length
11205: contig of 1561 bp in length
12766: gap of unknown length
12767: contig of 1849 bp in length
12867: gap of unknown length
14815: contig of 1058 bp in length
15873: contig of 1248 bp in length
15974: contig of 1124 bp in length
17221: gap of unknown length
17321: contig of 1283 bp in length
18445: gap of unknown length
18454: contig of 1283 bp in length
19828: gap of unknown length
21841: contig of 1913 bp in length
21941: gap of unknown length
23280: contig of 1339 bp in length
23281: gap of unknown length
23282: contig of 1841 bp in length
23283: gap of unknown length
23284: contig of 1588 bp in length
25009: contig of 1000 bp in length
27010: gap of unknown length
28109: contig of 1675 bp in length
29784: gap of unknown length
29884: contig of 1666 bp in length
31550: gap of unknown length
31650: contig of 1510 bp in length
33160: gap of unknown length
33260: contig of 1412 bp in length
34672: gap of unknown length
36257: contig of 1485 bp in length
36357: gap of unknown length
37408: contig of 1051 bp in length
37508: gap of unknown length
39418: contig of 1910 bp in length
39518: gap of unknown length
40833: contig of 1315 bp in length
40933: gap of unknown length
42346: contig of 1413 bp in length
42347: gap of unknown length
42447: contig of 1271 bp in length
4317: gap of unknown length
4318: contig of 1186 bp in length
45003: gap of unknown length
45103: contig of 1929 bp in length
45104: gap of unknown length
47032: contig of 1319 bp in length
47133: gap of unknown length
48451: contig of unknown length
48551: gap of unknown length
48552: contig of 2252 bp in length
48553: gap of unknown length
50904: contig of 1146 bp in length
52049: gap of unknown length
52149: contig of 1338 bp in length
53487: gap of unknown length
53587: contig of 1278 bp in length
54865: gap of unknown length
54965: contig of 1927 bp in length
56992: gap of unknown length
56993: contig of 1476 bp in length
58469: gap of unknown length
58568: contig of 1036 bp in length
59604: gap of unknown length
59704: contig of 1728 bp in length
61432: gap of unknown length
61433: contig of 1893 bp in length
63425: gap of unknown length
63426: contig of 1693 bp in length
63525: gap of unknown length
65218: contig of 1444 bp in length
65219: gap of unknown length
66762: contig of 1224 bp in length
66763: gap of unknown length
68086: contig of 1371 bp in length
68087: gap of unknown length
69557: contig of 1886 bp in length
69558: gap of unknown length
71543: contig of 1163 bp in length
71643: gap of unknown length
72806: contig of 1705 bp in length
72907: gap of unknown length
74612: contig of 1831 bp in length
74711: gap of unknown length
76542: contig of 2180 bp in length
76543: gap of unknown length
78822: contig of 1095 bp in length
78922: gap of unknown length
80017: contig of 2389 bp in length
80117: gap of unknown length
82706: contig of 2207 bp in length
82707: gap of unknown length
85013: contig of 1611 bp in length
85014: gap of unknown length
85113: contig of 124767;
85724: contig of 124767;
Query Match 3.9%; Score 42; DB 2; Length 124767;
Best Local Similarity 47.4%; Pred. No. 11;
Matches 126; Conservative 0; Mismatches 140; Indels 0; Gaps 0;
```

```
Qy 663 GGCTAAGAAATCAGAAAGGAGAGCCGAAAGCTTCTAAGTAATCAAGAAAAATCTTGGAC 722
Db 95477 GGCTAAGAAATCTGGAGTGTAAATAATAAGTAATAACCAACCAACCAATTAAT 95418
Qy 723 AATAGCTTTTCCATAGAGAAACCGTGGAGGAGGATCGATACCAATAAATTCGATTAA 782
Db 95417 AATAAATAAATAATTCCTCAAGATTTTCCCAATGAGAAGATTAATAAATTAAGATTA 95358
Qy 783 AAAGGGAGTGTTCACAGCCATATCAGTGGAGCAACCTTCCCTTTGATTCTTGATA 842
Db 95357 AAATAAATAATCTCTTAAGCCCAACATTAAGTCCCTGTTCTTAGGTTGATGATA 95298
Qy 843 ATGACAGAGATGACAAAGAGTAGTTTTTGAAGAAAAATCTTAAGATGATGATGAT 902
Db 95297 ACAGCAAAATTTGTTTTTCTACCTTCACATGACAGATACAAAGCCGTCATGAGTGTGT 95238
```

QY 903 GTCAGCTAAAGCTTCTCTCTTTA 928  
 Db 95237 CCCAAATACAGCTTCTCTTGTGTA 95212

## RESULT 37

MC445/c  
 LOCUS MC445 363 bp DNA linear BCT 10-AUG-1994

DEFINITION Mycoplasma capricolum DNA for CONTIG MC445.  
 ACCESSION 233336  
 VERSION 233336.1 GI:516259  
 KEYWORDS adenylsuccinate synthetase.  
 SOURCE Mycoplasma capricolum.

ORGANISM Bacteria: Firmicutes; Mollicutes; Entomoplasmatales;

Entomoplasmataceae.

1 (bases 1 to 363)

Bork.P., Ouzounis.C., Casari.G., Sander.C., Dolan.M. and

Gillevet.P.,

Exploring the Mycoplasma capricolum genome: A Parasite reveals it's

physiology

Unpublished

2 (bases 1 to 363)

Gillevet.P., Ally.A., Barton.F., Brenner.S.E., Clark-Whitehead.R.,  
 Dolan.M., Douglas.N., Hsu.E., Purzcki.M.S., Richter.B., Russo.S.,  
 Sartell.J., Smith.S.W., Wang.C., Williams.J. and Gilbert.W.

Direct Submission

Submitted (13-JUL-1994) Mycoplasma capricolum Genome Project,  
 Harvard University, Cambridge, MA 02138 Prokaryotic Genomes Project  
 Institute for Computational Sciences and Informatics George Mason  
 University, Fairfax, Virginia, USA, 22030-4444 E-mail:  
 gillevet@euranus.nchgr.nih.gov

3 (bases 1 to 363)

Bork.P., Ouzounis.C., Casari.G., Schneider.R., Sander.C., Dolan.M.,

Gilbert.W. and Gillevet.P.M.

Exploring the Mycoplasma capricolum genome: a minimal cell reveals

its Physiology

Mol Microbiol. 16 (5), 935-967 (1995)

96059641

7476192

NOTES:

The tga codon is read as Tryptophan in Mycoplasma capricolum Coding  
 sequences below were putatively identified by Peer Bork using the  
 program GeneQuiz at the EMBL.

EMBL, Meyerhofstr.1, 69012 Heidelberg, Germany

E-mail bork@embl-heidelberg.de

More than 870,477 bases have been sequenced by the genomic walking  
 technique and assembled into 372 non-overlapping contigs (Accession  
 numbers Z33005 to Z33376) covering 214,528 base pairs. The length  
 of the contigs varies from 63 base pairs to 2049 base pairs with an  
 average length of 658 base pairs. The data consists of 13,091 bases  
 (6%) from one pass fluorescent sequencing and 201,437 bases (94%)  
 produced by genomic walking. Of the latter, 155,805 bases (77%)  
 have multiple coverage on at least one strand and 112,621 bases  
 (56%) are covered on both strands. There is a total of 968  
 ambiguous nucleotide assignments in the data set (

0.5%). We have compared 8,868 bases of our data with Mycoplasma

capricolum sequences already stored in the public databases and

note less than 0.7% difference between the two data sets (

including ambiguous calls, insertions, deletions and mismatches).

Location/Qualifiers

1..363

## FEATURES

source

/organism="Mycoplasma capricolum"

/strain="ATCC 7343(Kid)"

/db\_xref="taxon:2095"

complement(<1..>363)

/note="ORF identified by homology to SwissProt entry

swiss|P30520|PURA\_HUMAN: Probability 3.8-33"

/codon\_start=3

/evidence=not\_experimental

/transl\_table=4

/codon=(seq:"tga",aa:Ttp)

/product="adenylsuccinate synthetase"

TITLE

Direct Submission

/protein\_id="CAA83842.1"  
 /db\_xref="GI:530493"  
 /db\_xref="SPTREMBL:Q49087"  
 /translation="YFNNWNNSCKTSKSVIGIVKYNTRVSGAMPTEIKTELANKLR  
 ERGREGSNTGKPRRIGWLDLVALKYAIRVGGIDQLFLFDVLDVTEEEKIKCTAYKL  
 DNOIHSIPANENFKDV"

BASE COUNT 112 a 64 c 36 g 150 t 1 others

Query Match 3.8%; Score 41.8; DB 1; Length 363;

Best Local Similarity 48.7%; Pred. No. 23;

Matches 146; Conservative 0; Mismatches 148; Indels 6; Gaps 1;

QY 606 AAACGGGATATAGAAAAATGTTTCTTAATTCGAAGAATAATTAAGCTGCTTGGTGGT 665

Db 333 AAACAAATTCAAAGTCATAGGATGTTTAAAGCATATAACTAGAGTTGGTTCAGGAG 274

QY 666 CTAAGAATCAGAAGAGACCAAGCTTCTAAGTAATCAAGAAAAAATCTTGGACAAT 725

Db 273 CTATGCCAACNGAAATTTAAACTTGAACCTTGTCTAAACTAAGAGAACGTTGGAGAGAAT 214

QY 726 AGCTTTTCCATAGAGAAAAACCGTGAGGAAGAGTCGATACCATATAAATCCGATTAAAA 785

Db 213 ATGGTTCTTAATACAGGAACCAAGAAAGTAAGTTGATTAGATTAGTTGCTTTAAAAAT 154

QY 786 AGGGAGTGCTTTACAAACAGCCATATCATGTGGCAACCTTCCCTTTGATTCTTTGATAATG 845

Db 153 ATGCAATT-----AGAGTTGGTGAATTTGATCAATTTATTTAACTTTATTTGATGCTAT 100

QY 846 ACAGAAGATGAACAAAGAGTAAGCTTTTTCAAAAATCTTAAAAATAGAATGCATTTGTGTC 905

Db 99 TAGATACAGAAGAAAAATTTAAATTTTGACAGCTTATAAATTAGATAATCAATATTATTC 40

## RESULT 38

AC024393/c

LOCUS AC024393

DEFINITION Homo sapiens chromosome 1 clone RP11-711C8 map 1, LOW-PASS SEQUENCE

SAMPLING.

ACCESSION AC024393.1 GI:7108190

VERSION AC024393

KEYWORDS HTG; HTGS\_PHASE0.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 76141)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 1, clone RP11-711C8

Unpublished

2 (bases 1 to 76141)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,

Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,

Choquel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,

DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,

Fenster,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Grant,G., Hagos,R., Hearford,A., Horton,L.,

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,

Klein,J., Landers,T., Larcocque,K., Lehoczy,J., Levine,R.,

Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,

McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J.,

Meneus,L., Mihova,T., Miranda,C., Menga,V., Morrow,J., Naylor,J.,

Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivat,T.M.,

Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,

Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,

Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,

Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,

Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and

Zody,M.

Fri Jan 10 12:01:26 2003

us-09-438-185a-1\_1199590\_1200675.rge

Page 43

JOURNAL  
COMMENT

Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L6722

Center clone name: 711\_C\_8

-----

\* NOTE: This record contains 87 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1  
\* 736 835: gap of 100 bp  
\* 836 1595: contig of 760 bp in length  
\* 1596 1695: gap of 100 bp  
\* 1696 2461: contig of 766 bp in length  
\* 2462 2561: gap of 100 bp  
\* 2562 3364: contig of 803 bp in length  
\* 3365 3464: gap of 100 bp  
\* 3465 4249: contig of 785 bp in length  
\* 4250 4349: gap of 100 bp  
\* 4350 5141: contig of 792 bp in length  
\* 5142 5241: gap of 100 bp  
\* 5242 6017: contig of 776 bp in length  
\* 6018 6117: gap of 100 bp  
\* 6118 6900: contig of 783 bp in length  
\* 6901 7000: gap of 100 bp  
\* 7001 7784: contig of 784 bp in length  
\* 7785 7884: gap of 100 bp  
\* 7885 8666: contig of 782 bp in length  
\* 8667 8766: gap of 100 bp  
\* 8767 9542: contig of 776 bp in length  
\* 9543 9642: gap of 100 bp  
\* 9643 10391: contig of 749 bp in length  
\* 10392 10491: gap of 100 bp  
\* 10492 11278: contig of 787 bp in length  
\* 11279 11378: gap of 100 bp  
\* 11379 12148: contig of 770 bp in length  
\* 12149 12248: gap of 100 bp  
\* 12249 13033: contig of 785 bp in length  
\* 13034 13133: gap of 100 bp  
\* 13134 13889: contig of 756 bp in length  
\* 13890 13989: gap of 100 bp  
\* 13990 14814: contig of 825 bp in length  
\* 14815 14914: gap of 100 bp  
\* 14915 15682: contig of 768 bp in length  
\* 15683 15782: gap of 100 bp  
\* 15783 16561: contig of 779 bp in length  
\* 16562 16661: gap of 100 bp  
\* 16662 17452: contig of 791 bp in length  
\* 17453 17552: gap of 100 bp  
\* 17553 18316: contig of 764 bp in length  
\* 18317 18416: gap of 100 bp  
\* 18417 19192: contig of 776 bp in length  
\* 19193 19292: gap of 100 bp  
\* 19293 20055: contig of 763 bp in length  
\* 20056 20155: gap of 100 bp  
\* 20156 20919: contig of 764 bp in length  
\* 20920 21019: gap of 100 bp

\* 21020 21785: contig of 766 bp in length  
\* 21786 21885: gap of 100 bp  
\* 21886 22661: contig of 776 bp in length  
\* 22662 22761: gap of 100 bp  
\* 22762 23543: contig of 782 bp in length  
\* 23544 23643: gap of 100 bp  
\* 23644 24426: contig of 783 bp in length  
\* 24427 24526: gap of 100 bp  
\* 24527 25297: contig of 771 bp in length  
\* 25298 25397: gap of 100 bp  
\* 25398 26196: contig of 799 bp in length  
\* 26197 26296: gap of 100 bp  
\* 26297 27086: contig of 790 bp in length  
\* 27087 27186: gap of 100 bp  
\* 27187 27978: contig of 792 bp in length  
\* 27979 28078: gap of 100 bp  
\* 28079 28864: contig of 786 bp in length  
\* 28865 28964: gap of 100 bp  
\* 28965 29718: contig of 754 bp in length  
\* 29719 29818: gap of 100 bp  
\* 29819 30601: contig of 783 bp in length  
\* 30602 30701: gap of 100 bp  
\* 30702 31467: contig of 766 bp in length  
\* 31468 31567: gap of 100 bp  
\* 31568 32343: contig of 776 bp in length  
\* 32344 32443: gap of 100 bp  
\* 32444 33233: contig of 790 bp in length  
\* 33234 33333: gap of 100 bp  
\* 33334 34108: contig of 775 bp in length  
\* 34109 34208: gap of 100 bp  
\* 34209 34982: contig of 774 bp in length  
\* 34983 35082: gap of 100 bp  
\* 35083 35863: contig of 781 bp in length  
\* 35864 36716: contig of 753 bp in length  
\* 36717 36816: gap of 100 bp  
\* 36817 37591: contig of 775 bp in length  
\* 37592 37691: gap of 100 bp  
\* 37692 38467: contig of 776 bp in length  
\* 38468 39336: contig of 769 bp in length  
\* 39337 39436: gap of 100 bp  
\* 39437 40184: contig of 748 bp in length  
\* 40185 40284: gap of 100 bp  
\* 40285 41091: contig of 807 bp in length  
\* 41092 41191: gap of 100 bp  
\* 41192 41942: contig of 751 bp in length  
\* 41943 42042: gap of 100 bp  
\* 42043 42823: contig of 781 bp in length  
\* 42824 42923: gap of 100 bp  
\* 42924 43708: contig of 785 bp in length  
\* 43709 43808: gap of 100 bp  
\* 43809 44580: contig of 772 bp in length  
\* 44581 44680: gap of 100 bp  
\* 44681 45425: contig of 745 bp in length  
\* 45426 45525: gap of 100 bp  
\* 45526 46309: contig of 784 bp in length  
\* 46310 46409: gap of 100 bp  
\* 46410 47193: contig of 784 bp in length  
\* 47194 47293: gap of 100 bp  
\* 47294 48062: contig of 773 bp in length  
\* 48067 48166: gap of 100 bp  
\* 48167 48931: contig of 765 bp in length  
\* 48932 49031: gap of 100 bp  
\* 49032 49812: contig of 781 bp in length  
\* 49813 49912: gap of 100 bp  
\* 49913 50664: contig of 752 bp in length  
\* 50665 50764: gap of 100 bp  
\* 50765 51522: contig of 758 bp in length  
\* 51523 51622: gap of 100 bp  
\* 51623 52382: contig of 760 bp in length  
\* 52383 52482: gap of 100 bp  
\* 52483 53264: contig of 782 bp in length

```

* 53265 53364: gap of 100 bp
* 53365 54144: contig of 780 bp in length
* 54145 54244: gap of 100 bp
* 54245 55007: contig of 763 bp in length
* 55008 55107: gap of 100 bp
* 55108 55883: contig of 776 bp in length
* 55884 55983: gap of 100 bp
* 55984 56757: contig of 774 bp in length
* 56758 56857: gap of 100 bp
* 56858 57632: contig of 775 bp in length
* 57633 57732: gap of 100 bp
* 57733 58512: contig of 780 bp in length
* 58513 58612: gap of 100 bp
* 58613 59370: contig of 758 bp in length
* 59371 59470: gap of 100 bp
* 59471 60240: contig of 770 bp in length
* 60241 60340: gap of 100 bp

Query Match      3.8%; Score 41.8; DB 2; Length 76141;
Best Local Similarity 46.3%; Pred. No. 13;
Matches 100; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 596 ACAATCCTGAAACGGGATAATAGAAATAATGCTTCTTAATTCGAGAAATTTAATGACT 655
DB 40996 AAGATATTTAAATCTTATAAAGAAAAATAATAATATTATAATAATAATTTAATAAT 40937

QY 656 GCTTGTGGTCTTAAGAAATCAGAAAGAGACCGGAAAGCTTCTTAAGTAATCAAGAAAAAT 715
DB 40936 TATTAATAAAAAAANNAAAAAATTAATAAATAAANNNTTAAATAAANNAGAAA 40877

QY 716 CTTCGACAAATAGCTTTTCCATAGACAAACCGTGCAGAAAGGATCGATACCAATAATTC 775
DB 40876 ATANTAGATNTNNNTGGGTTTAAAAAATAAAGAAATAAAAAAATAAAAAAATAAT 40817

QY 776 CGATTAATAAAGGAGTGCTTTACACACCCATATCA 811
DB 40816 NAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 40781

RESULT 39
AC105744
LOCUS
DEFINITION
Oryza sativa chromosome 10 clone OSJNB0009C07, HTG 25-APR-2002
PROGRESS ***, 8 ordered pieces.
AC105744
VERSION
KEYWORDS
HTG; HTGS_PHASE2.
SOURCE
Oryza sativa.
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 115666)
Buell, C.R., Yuan, Q., Ouyang, S., Liu, J., Tallon, L., Gansberger, K.,
Kim, M., Overton II, L., Bera, J., Tsitrin, T., Krol, M., Jarrahi, B.,
Jin, S., Koo, H., Zisman, V., Hsiao, J., Blunt, S., Vanaken, S.,
Uterbach, T., Feldblyum, T., Yang, Q., Haas, B., Suh, B., Peterson, J.,
Quackenbush, J., White, O., Salzberg, S. and Fraser, C.
Oryza sativa ssp. japonica cv. Nipponbare OSJNB0009C07 BAC genomic
sequence
Unpublished
2 (bases 1 to 115666)
Buell, R.
Direct Submission
Submitted (09-JAN-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
3 (bases 1 to 115666)
Buell, R.
Direct Submission
Submitted (25-APR-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
JOURNAL
COMMENT
On Apr 25, 2002 this sequence version replaced gi:20153307.
* NOTE: This is a 'working draft' sequence. It currently

```

```

* consists of 8 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 18477: contig of 18477 bp in length
* 18478 18578: gap of unknown length
* 18578 20885: contig of 2308 bp in length
* 20886 20986: gap of unknown length
* 20986 25966: contig of 4981 bp in length
* 25967 26067: gap of unknown length
* 26067 49469: contig of 23403 bp in length
* 49470 52759: gap of unknown length
* 52759 52659: contig of 3090 bp in length
* 52660 82861: contig of 30102 bp in length
* 82862 82961: gap of unknown length
* 82962 110436: contig of 27475 bp in length
* 110437 110536: gap of unknown length
* 110537 115666: contig of 5130 bp in length.
FEATURES
Location/Qualifiers
source
1..115666
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="10"
/clone="OSJNB0009C07"
/notes="japonica cultivar-group"
BASE COUNT 32490 a 24957 c 25425 g 31751 t 1043 others
ORIGIN

Query Match      3.8%; Score 41.8; DB 2; Length 115666;
Best Local Similarity 12.3%; Pred. No. 12;
Matches 37; Conservative 136; Mismatches 128; Indels 0; Gaps 0;

QY 442 AGAAATGAGGGATGTAGAACCAAGGACGTCCTCAAAAGTCATGATGATCAGCAG 501
DB 46372 RKGNGKKGSGSSGCGYKMGSGRKGSRKMSRRRGRTKGSYKWSWSWS 46431

QY 502 AGTTAAGAGAAATATCTTATCGAGAGTTCGCATTACAGAGCAATGGGAAAGTAACG 561
DB 46432 RKWGMNTSYMMYTCRRWSMMYSMRWMMYMKMSYRKYSKKAMKSKAGWSK 46491

QY 562 GTCTGCAACAGACAGACAAGTATTGATGGGGAGCTACAATCTCAAAACGGATAATAGGA 621
DB 46492 KRCKSKWKAKSRMVKMTYWKWSSSCHWRTWYWKCCMKMKWMTAWKWM 46551

QY 622 AAATGTGTTTCTAATTCGAGAAATTTAATGACTCTTGATGCTTAAGAAATCAGAAAG 581
DB 46552 WMRWKSMMKATARRMWMWRCMGAWRGRRRSAGRRMRCSSNSMKMKKASWAGS 46611

QY 682 GAGACGAAAGCTTCTTAAGTAATCAAGAAAAAATCTTGACAAATAGCTTTTCCATAGAGA 741
DB 46612 SACMVCYMYACACSRWYMMKMAAAYAGARWKWMAAGAGAGWAGTGTGTTAGTAGTRA 46671

QY 742 A 742
DB 46672 A 46672

RESULT 40
AL391883
LOCUS
DEFINITION
Human DNA sequence from clone RP11-91K11 on chromosome 1, complete
sequence.
ACCESSION
AL391883.16
VERSION
AL391883.16
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. Ellington, A. Direct Submission

Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

On Nov 16, 2001 this sequence version replaced gi:15131273. During difference assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at: <http://www.sanger.ac.uk/projects/C-elegans/wormpep>

This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at: <http://www.sanger.ac.uk/HGP/Chr1>

RP11-91K11 is from the library RP11-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-91K11. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true right end of clone RP11-91K11 is at 126371 in this sequence. The true left end of clone RP11-460G22 is at 31870 in this sequence. The true right end of clone RP1-301K23 is at 2000 in this sequence.

FEATURES  
source  
1. 126371 Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-91K11"  
/clone\_lib="RP11-11.1"  
BASE COUNT 35354 a 27351 c 28154 g 35512 t

Query Match 3.8%; Score 41.8; DB 9; Length 126371;  
Best Local Similarity 47.8%; Pred. No. 12; Indels 0; Gaps 0;  
Matches 121; Conservative 0; Mismatches 132

QY 698 AGTAATCAAGAAAAATCTTGGACATAGCTTTCCATAGAGAAACCGTGAGGAAGG 757  
DB 45012 AAGGAATATAACAAATATTTGTAATAGTATGACACATCTAAGAAATTTAAGG 45071  
QY 758 AGTCGATACCAATAATCCGATTTAAAAAGGGAGTCTTACCAACAGCCATVACGTCGAG 817  
DB 45072 AGAAGTGATTCGGTATCTGCAATTTTCTTGAATTTACATATAAAAAAATAGATGAT 45131  
QY 818 CAACCTTCCTCTTTGATCTTTGATATGACAGAGATGACAAAGAGTAACTTTTGGAA 877  
DB 45132 GAATGGTTAGAGGATGATAGTAAAGCAATGAAGCAAGTATATAGAAATATTGATGTA 45191  
QY 878 AATTCTAAATAGAAATGCATTTGTGTCGAGCTTAAAGCTTGTCTCTTTATTTTCCTTT 937  
DB 45192 GAATCTAGACGGTGGATATATTGGTATGTACCATAACTTCTCTCTCTCTCTCTTTTA 45251  
QY 938 TGTAGATGATTC 950

DB 45252 TACACAGTCTTG 45264

RESULT 41  
AC004153 156060 bp DNA linear HTG 12-AUG-2000  
LOCUS Plasmodium falciparum chromosome 12 clone 307, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\*, 2 unordered pieces.  
AC004153  
AC004153 7 GI:97977733  
VERSION HTG; HTGS\_PHASE1.  
KEYWORDS Plasmodium falciparum.  
SOURCE Plasmodium falciparum  
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
REFERENCE 1 (bases 1 to 156060)  
Hyman, R.W., Fung, E.L., Qin, F., Rowley, D., Mao, J., Tamaki, T.,  
Kurd, O.B., Conway, A.B. and Davis, R.W.  
Plasmodium falciparum 307 chromosome 12  
Unpublished  
2 (bases 1 to 156060)  
Hyman, R.W., Qin, F., Fung, E.L., Conway, A.B. and Davis, R.W.  
Direct Submission  
Submitted (18-FEB-1998) Stanford DNA Sequencing and Technology  
Center, Stanford University, 855 California Avenue, Palo Alto, CA  
94304, USA

COMMENT  
On Aug 12, 2000 this sequence version replaced gi:8810454.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 102169: contig of 102169 bp in length  
\* 102170 102369: gap of unknown length  
\* 102370 156060: contig of 53691 bp in length.

FEATURES  
source  
1. 156060 Location/Qualifiers  
/organism="Plasmodium falciparum"  
/db\_xref="taxon:5833"  
/chromosome="12"  
/clone="PFYAC812"  
/clone="3D7"  
BASE COUNT 62615 a 14889 c 15137 g 63219 t 200 others

ORIGIN  
Query Match 3.8%; Score 41.8; DB 2; Length 156060;  
Best Local Similarity 45.3%; Pred. No. 12;  
Matches 151; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

QY 568 CAACAGACACAAGTATTGATGGGAGCTACAAATCTGAAACGGGATAATAGGAAATG 627  
DB 110223 CAAAAAATAAATAATGATGATAATTCATATGGAATGGGAAATTTCAATCCAC 110282  
QY 628 TGTCTTAAATCGAAGAAATTTAATGACTGCTGATGCTTAAGAATCAGAAAGGAGACC 687  
DB 110283 AATTACATATTTCAATACCTTTTGAACCTAATGAGATATGTTCTTTCTAAAGAAAAAC 110342  
QY 688 GAAAGCTTCTAAGTAATCAAGAAAAATCTTGGACATAAGCTTTTCCATAGAGAAAAACCG 747  
DB 110343 ATGATATCTCAATGATGGAGAAAAAGGGTTCTAATAATATGCAAAAAAACAATCCA 110402  
QY 748 TGAGGAAGGAGTCGATACCAATAATTCGATTTAAAAAGGGAGTCTTACAACAGCAT 807  
DB 110403 AAGAAAAAGTCTAGATAATACAAATGATATAGAAACATATAATAAATAACATTA 110462  
QY 808 ATCAGTGGAGCAACCTTCCCTTTTGAATTTGATTTGATATGACAGAAGTCAACAAGATAA 867  
DB 110463 ATAATAAAGAAAAAGAAATGATGCTTTTATCAATGAAATTCATATATAAATA 110522  
QY 868 GTTTTGAATAATCTTAAATAAGAAATGCATTT 900





Db 13383 TACACACTCTTGC 13395

## RESULT 44

AC109165/c

LOCUS

DEFINITION

pieces.

AC109165

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

TITLES

JOURNAL

REFERENCE

AUTHORS

AC109165 200279 bp DNA linear HTG 25-APR-2002  
Mus musculus clone RP23-30B3, WORKING DRAFT SEQUENCE, 16 ordered  
pieces.

AC109165 3 GI:20303726

HTG: HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

Mus musculus.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 200279)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Unpublished

2 (bases 1 to 200279)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,

Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,

Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,

Choepe,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,

Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,

Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,

Finde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,

Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,

Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,

Landers,T., Lehoczyk,J., Levine,R., Liu,G., MacClean,C.,

MacDonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,

McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T.,

Mleung,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,

Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,

Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,

Retta,B., Riback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,

Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S.,

Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,

Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,

Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,

Zainoun,J., Zemek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (03-FEB-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 200279)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,

Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,

Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,

Chazaro,B., Choepe,Y., Colangelo,M., Collins,S., Collamore,A.,

Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,

Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,

Finde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,

Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,

Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,

Landers,T., Lehoczyk,J., Levine,R., Lindblad-Toh,K., Liu,G.,

MacClean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,

McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,

Mihova,T., Mleung,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,

Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,

Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,

Raymond,C., Retta,B., Riback,M., Riley,R., Rise,C., Rogov,P.,

Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,

Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,

Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,

Zainoun,J., Zemek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (25-APR-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Apr 25, 2002 this sequence version replaced gi:20258472.

All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center Project name: L18963

Center Clone name: 30\_B.3

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 19597 bases at least Q40

Consensus quality: 197569 bases at least Q30

Consensus quality: 198340 bases at least Q20

Insert size: 198000; agarose-fp

Quality coverage: 7.0 in Q20 bases; agarose-fp

Quality coverage: 7.0 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 16 contigs. Gaps between the contigs

\* are represented as runs of N. The order of the pieces

\* is believed to be correct as given, however the sizes

\* of the gaps between them are based on estimates that have

\* provided by the submitter.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

\* 1 564: contig of 564 bp in length

\* 565 564: gap of 100 bp

\* 566 1284: contig of 620 bp in length

\* 1385 1384: gap of 100 bp

\* 1385 2041: contig of 637 bp in length

\* 2042 2141: gap of 100 bp

\* 2142 3109: contig of 968 bp in length

\* 3110 3209: gap of 100 bp

\* 3210 4516: contig of 1307 bp in length

\* 4517 4616: gap of 100 bp

\* 4617 6565: contig of 1949 bp in length

\* 6566 6665: gap of 100 bp

\* 6666 9499: contig of 2834 bp in length

\* 9500 9599: gap of 100 bp

\* 9600 13090: contig of 3491 bp in length

\* 13091 13190: gap of 100 bp

\* 13191 17101: contig of 3911 bp in length

\* 17102 17201: gap of 100 bp

\* 17202 22059: contig of 4858 bp in length

\* 22060 22159: gap of 100 bp

\* 22160 30887: contig of 8728 bp in length

\* 30888 30987: gap of 100 bp

\* 30988 42016: contig of 11029 bp in length

\* 42017 42116: gap of 100 bp

\* 42117 54621: contig of 12505 bp in length

\* 54622 54721: gap of 100 bp

\* 54722 80967: contig of 26246 bp in length

\* 80968 81067: gap of 100 bp

\* 81068 136318: contig of 55251 bp in length

\* 136319 136418: gap of 100 bp

\* 136419 200279: contig of 63861 bp in length.

\* Location/Qualifiers

1. 200279

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="RP23-30B3"

/clone\_lib="RPCI-23 Female Mouse BAC"

1. 564

/note="assembly\_fragment"

665 1284

/note="assembly\_fragment"

FEATURES  
source

misc\_feature

misc\_feature











TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Direct Submission  
Unpublished  
2 (bases 1 to 103337)  
Worley, K.C.  
Direct Submission  
Submitted (19-MAR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 103337)  
Worley, K.C.  
Direct Submission  
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 12, 2002 this sequence version replaced gi:19549112.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GNZA  
Center clone name: CH230-74P14  
----- Summary Statistics  
Sequencing vector: Plasmid  
Chemistry: Dye-terminator Big Dye: 100% of reads  
Assembly program: phrap version 0.990329  
Consensus quality: 50038 bases at least Q40  
Consensus quality: 53149 bases at least Q30  
Consensus quality: 56140 bases at least Q20  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
consists of 50 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1068: contig of 1068 bp in length  
1168: gap of unknown length  
1169: contig of 1165 bp in length  
2333: gap of unknown length  
2434: contig of 1010 bp in length  
3443: gap of unknown length  
3543: contig of 1012 bp in length  
4555: gap of unknown length  
4635: contig of 1492 bp in length  
6147: gap of unknown length  
6247: gap of unknown length  
7409: contig of 1162 bp in length  
9245: contig of 1736 bp in length  
9345: gap of unknown length  
10347: contig of 1002 bp in length  
10348: gap of unknown length  
10448: contig of 1342 bp in length  
11789: gap of unknown length  
11790: contig of 1075 bp in length  
11890: contig of 1075 bp in length  
12964: gap of unknown length  
13064: contig of 1531 bp in length  
13065: gap of unknown length  
14595: gap of unknown length  
14596: contig of 1468 bp in length  
16163: gap of unknown length  
16264: contig of 1156 bp in length  
17419: gap of unknown length  
17519: contig of 1515 bp in length  
19034: contig of 1515 bp in length  
19035: gap of unknown length  
20254: contig of 1120 bp in length  
20354: gap of unknown length  
22053: contig of 1699 bp in length  
22153: gap of unknown length

22154: contig of 1061 bp in length  
23215: gap of unknown length  
23315: contig of 1570 bp in length  
24884: gap of unknown length  
24885: contig of 1649 bp in length  
26633: gap of unknown length  
26733: contig of 1831 bp in length  
28734: gap of unknown length  
28664: contig of 1903 bp in length  
30567: gap of unknown length  
30568: contig of 1451 bp in length  
32118: gap of unknown length  
32119: contig of 1430 bp in length  
33648: gap of unknown length  
33748: contig of 1575 bp in length  
35233: gap of unknown length  
35423: contig of 1571 bp in length  
36994: gap of unknown length  
37094: contig of 2498 bp in length  
39592: gap of unknown length  
39692: contig of 1308 bp in length  
41000: gap of unknown length  
41001: contig of 1422 bp in length  
42522: gap of unknown length  
42622: contig of 1896 bp in length  
44518: gap of unknown length  
44618: contig of 1426 bp in length  
46044: gap of unknown length  
46144: contig of 1794 bp in length  
47938: gap of unknown length  
48038: contig of 1627 bp in length  
49665: gap of unknown length  
49765: contig of 1632 bp in length  
51397: gap of unknown length  
51497: contig of 1155 bp in length  
52552: gap of unknown length  
52553: contig of 1891 bp in length  
54543: gap of unknown length  
54743: contig of 2505 bp in length  
57248: gap of unknown length  
57348: contig of 1815 bp in length  
59163: gap of unknown length  
59263: contig of 1500 bp in length  
60763: gap of unknown length  
60863: contig of 1090 bp in length  
61953: gap of unknown length  
62053: contig of 2215 bp in length  
64268: gap of unknown length  
64368: contig of 1913 bp in length  
66281: gap of unknown length  
66381: contig of 2184 bp in length  
68565: gap of unknown length  
68665: contig of 2990 bp in length  
71655: gap of unknown length  
71755: contig of 1444 bp in length  
73299: gap of unknown length  
73300: contig of 3424 bp in length  
76723: gap of unknown length  
76824: contig of 3676 bp in length  
80489: gap of unknown length  
80500: contig of 4329 bp in length  
85128: gap of unknown length  
85229: contig of 6390 bp in length  
91618: gap of unknown length  
91718: contig of 5497 bp in length  
97216: gap of unknown length  
97315: contig of 6022 bp in length.  
97316: 103337  
Location/Qualifiers  
1. 103337  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-74P14"

30735 a 18325 c 18083 g 29634 t 6560 others

BASE COUNT

## ORIGIN

```
Query Match      3.8%; Score 41.6; DB 2; Length 103337;
Best Local Similarity 51.6%; Pred. No. 14;
Matches 95; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 678 AAGGACCGAAGAGCTTCTAAGTAATCAAGAAAAATCTTGGACAAATAGCTTTTCCATA 737
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 57477 AAGAAAAGCTGCTTTTGGAGAGATTAACAAGATAGATAAAACCCCTTAGCAAACTATC 57536
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 738 GAGAAAACCGTGAGGAAGAGCTGCAATACCAATATCCGATTAAAAAGGAGGCTGCTTA 797
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 57537 TAAAGGCTCCAGAGGAAATTTCCCAATTCAGAAAATCAGATATGAAGGTAGAGTAG 57596
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 798 CAACAGCCATATCAGTGGAGCAACCTTCCCTTGTATCTTTGATATGACAGAGATGAA 857
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 57597 CACTGTATAGAGGAGTTCAAAANAATCATCATCTCTACTACAAAAGCTTATCTAGAA 57656
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 858 CAAA 861
Db 57657 CAAA 57660

RESULT 48
AL353694_1
WPCOMMENT
Sequence split into 4 fragments LOCUS AL353694 Accession AL353694
Fragment Name      Begin      End
AL353694.0         1      110000
AL353694.1        100001      210000
AL353694.2        200001      310000
AL353694.3        300001      356330
Continuation (2 of 4) of AL353694 from base 100001 (AL353694 Homo sapiens chromosome 6)

Query Match      3.8%; Score 41.6; DB 2; Length 110000;
Best Local Similarity 54.6%; Pred. No. 14;
Matches 83; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 659 AGAATCAGAAAGGAGCCGAAAGCTTCTAAGTAATCAAGAAAAATCTTGGACAAATAGC 728
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 91714 ACAAACATCAATGAACAAAAAATTTGTTTTTGAAGATATAAATTTATGCAACCA 91773
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 729 TTTTCCATAGAGAAACCGTGAGGAAGAGCTGATACCAATAATCCGATTAAAAAGG 788
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 91774 TTAGCCAGACAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 91833
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 789 GAGTGCCTACACAGCCATATCAGTGGAGCAA 820
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 91834 GAGACATTACAACCTGATACCCACAGAAATTCAA 91865
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 49
AL354818/c
LOCUS
DEFINITION
Homo sapiens chromosome 13 clone RP11-78L16, ** SEQUENCING IN
PROGRESS ***, 25 unordered pieces.
ACCESSION
AL354818
VERSION
AL354818.3 GI:10039795
KEYWORDS
HTG; HTGS-PHASE1; HTGS-CANCELLED.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Direct Submission
Burtion.J.
Submitted (10-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Sep 9, 2000 this sequence version replaced gi:9581627.
-----
Center: Sanger Centre
Genome Center
Center code: SC
```

Web site: <http://www.sanger.ac.uk>Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)

----- Project Information

Center project name: BA78L16

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 13724 bases at least Q40

Consensus quality: 139166 bases at least Q30

Consensus quality: 142544 bases at least Q20

Insert size: 146149; sum-of-contigs

Insert size: 153445; 5% error; agarose-fp

Quality coverage: 2.84x in Q20 bases; sum-of-contigs Quality

coverage: 3.100x in Q20 bases; agarose-fp

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 25 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 4894: contig of 4894 bp in length

\* 4895 4994: gap of 100 bp

\* 4995 16378: contig of 11384 bp in length

\* 16379 16478: gap of 100 bp

\* 16479 18490: contig of 2012 bp in length

\* 18491 18590: gap of 100 bp

\* 18591 21846: contig of 3256 bp in length

\* 21847 21946: gap of 100 bp

\* 21947 26998: contig of 5052 bp in length

\* 26999 27098: gap of 100 bp

\* 27099 29374: contig of 2276 bp in length

\* 29375 29474: gap of 100 bp

\* 29475 32601: contig of 3127 bp in length

\* 32602 32701: gap of 100 bp

\* 32702 38362: contig of 5661 bp in length

\* 38363 38462: gap of 100 bp

\* 38463 48272: contig of 9810 bp in length

\* 48273 48372: gap of 100 bp

\* 48373 52389: contig of 4017 bp in length

\* 52390 52489: gap of 100 bp

\* 52490 62396: contig of 9907 bp in length

\* 62397 62496: gap of 100 bp

\* 62497 67828: contig of 5332 bp in length

\* 67829 67928: gap of 100 bp

\* 67929 70819: contig of 2891 bp in length

\* 70820 70919: gap of 100 bp

\* 70920 79571: contig of 8652 bp in length

\* 79572 79671: gap of 100 bp

\* 79672 83474: contig of 3803 bp in length

\* 83475 83574: gap of 100 bp

\* 83575 90030: contig of 6456 bp in length

\* 90031 90130: gap of 100 bp

\* 90131 112698: contig of 22568 bp in length

\* 112699 112798: gap of 100 bp

\* 112799 118187: contig of 5389 bp in length

\* 118188 118287: gap of 100 bp

\* 118288 122327: contig of 4040 bp in length

\* 122328 122427: gap of 100 bp

\* 122428 130709: contig of 8282 bp in length

\* 130710 130809: gap of 100 bp

\* 130810 132827: contig of 2018 bp in length

\* 132828 132927: gap of 100 bp

\* 132928 136248: contig of 3321 bp in length

\* 136249 136348: gap of 100 bp

\* 136349 139531: contig of 3183 bp in length

\* 139532 139631: gap of 100 bp

\* 139632 143470: contig of 3839 bp in length

\* 143471 143570: gap of 100 bp

\* 143571 148549: contig of 4979 bp in length.



```

http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8718
Center clone name: 465_J_17
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 203149 bases at least Q40
Consensus quality: 209316 bases at least Q30
Consensus quality: 211902 bases at least Q20
Insert size: 220000; agarose-fp
Insert size: 213296; sum-of-contigs
Quality coverage: 4.3 in Q20 bases; agarose-fp
Quality coverage: 4.4 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1011: contig of 1011 bp in length
* 1012 1111: gap of 100 bp
* 1112 2633: contig of 1522 bp in length
* 2634 2733: gap of 100 bp
* 2734 4093: contig of 1360 bp in length
* 4094 4193: gap of 100 bp
* 4194 5298: contig of 1105 bp in length
* 5299 5398: gap of 100 bp
* 5399 7088: contig of 1690 bp in length
* 7089 7188: gap of 100 bp
* 7189 7889: contig of 701 bp in length
* 7890 7989: gap of 100 bp
* 7990 10304: contig of 2315 bp in length
* 10305 10404: gap of 100 bp
* 10405 12953: contig of 2551 bp in length
* 12956 13053: gap of 100 bp
* 13056 16426: contig of 3371 bp in length
* 16427 16526: gap of 100 bp
* 16527 19178: contig of 2652 bp in length
* 19179 19278: gap of 100 bp
* 19279 23519: contig of 4241 bp in length
* 23520 26853: contig of 3234 bp in length
* 26854 26954: gap of 100 bp
* 26954 31292: contig of 4339 bp in length
* 31293 31392: gap of 100 bp
* 31393 36401: contig of 5009 bp in length
* 36402 36501: gap of 100 bp
* 36502 41321: contig of 4820 bp in length
* 41322 41421: gap of 100 bp
* 41422 47286: contig of 5865 bp in length
* 47287 47386: gap of 100 bp
* 47387 54497: contig of 7111 bp in length
* 54498 54597: gap of 100 bp
* 54598 60757: contig of 6160 bp in length
* 60758 60857: gap of 100 bp
* 60858 68004: contig of 7147 bp in length
* 68005 68104: gap of 100 bp
* 68105 77167: contig of 9063 bp in length
* 77168 77267: gap of 100 bp
* 77268 89567: contig of 11300 bp in length
* 89568 89667: gap of 100 bp
* 89668 101419: contig of 12752 bp in length
* 101420 101519: gap of 100 bp

```

---

```

* 101520 114909: contig of 13390 bp in length
* 114910 115009: gap of 100 bp
* 115010 130182: contig of 15173 bp in length
* 130183 130282: gap of 100 bp
* 130283 151503: contig of 21221 bp in length
* 151504 151603: gap of 100 bp
* 151604 179979: contig of 28376 bp in length
* 179980 180079: gap of 100 bp
* 180080 215896: contig of 35817 bp in length.
FEATURES
    Location/Qualifiers
        1..215896
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="15"
            /map="15"
            /clone="RP11-465J17"
            /clone_lib="RPC1-11 Human Male BAC"
        1..1011
            misc_feature
            note="assembly_fragment"
        1112..2633
            misc_feature
            note="assembly_fragment"
        2734..4093
            misc_feature
            note="assembly_fragment"
        4194..5298
            misc_feature
            note="assembly_fragment"
        5399..7088
            misc_feature
            note="assembly_fragment"
        7189..7889
            misc_feature
            note="assembly_fragment"
            clone_end="7"
            vector_side="right"
        7990..10304
            misc_feature
            note="assembly_fragment"
        10405..12955
            misc_feature
            note="assembly_fragment"
        13056..16426
            misc_feature
            note="assembly_fragment"
        16527..19178
            misc_feature
            note="assembly_fragment"
        19279..23519
            misc_feature
            note="assembly_fragment"
        23620..26853
            misc_feature
            note="assembly_fragment"
        26954..31292
            misc_feature
            note="assembly_fragment"
        31393..36401
            misc_feature
            note="assembly_fragment"
        36502..41321
            misc_feature
            note="assembly_fragment"
        41422..47286
            misc_feature
            note="assembly_fragment"
        47387..54497
            misc_feature
            note="assembly_fragment"
        54598..60757
            misc_feature
            note="assembly_fragment"
            clone_end="SP6"
            vector_side="left"
        60858..68004
            misc_feature
            note="assembly_fragment"
        68105..77167
            misc_feature
            note="assembly_fragment"
        77268..89567
            misc_feature
            note="assembly_fragment"
        89668..101419
            misc_feature
            note="assembly_fragment"
        101520..114909
            misc_feature
            note="assembly_fragment"
        115010..130182
            misc_feature
            note="assembly_fragment"
        130283..151503
            misc_feature
            note="assembly_fragment"
        151604..179979
            misc_feature
            note="assembly_fragment"

```



```
misc_feature 180080..215896
BASE COUNT 66799 a 38601 c 40541 g 67354 t 2601 others
ORIGIN
Query Match 3.8%; Score 41.6; DB 2; Length 215896;
Best Local Similarity 48.6%; Pred. No. 13;
Matches 142; Conservative 0; Mismatches 149; Indels 1; Gaps 1;
QY 598 AAATCCTGAAACGGGATAATAGGAAAAATGTGTTCTTAATTTCGAGAGAAATTTAATGACTGC 657
Db 113768 AAAGATTTAACTGGGATATTCAGATTAAAGATAAAAAACCTTATGCTATTTTCATAGA 113827
QY 658 TTGATGGTCTAAGAAATCAGAAAGGAGACCGAAAGCTTCTAAGTAATCAAGAAAAAATCT 717
Db 113828 AGGATGAATGGAACACACAGACATTATTTTAAGAGTATTAAAGGCTCTTATACATATGTT 113887
QY 718 TGGACAATAGCTTTTCCATAGAGAAAAACCGTGAGGAAAGGAGTCGATACGATAAATTCGG 777
Db 113888 TACATATCAAAATTTATCTATAATATCCAGTGTGAACGTGAGTCTCTAAGATTAATTCCT 113947
QY 778 ATTAAAAAGGGAGTCTTTACACACGCCATATCAGTGGAGCAACCTTCCTTTGATTCCT 837
Db 113948 GCTCTACTACTGTATAATATATAAANAACCTTTAATTAGRCNAGCATCTTTAGAGGCTA 114007
QY 838 T-GATAATGACAGAAAGTGAACAAGAGTAAGTTTTCGAAAAATTCFAAAT 888
Db 114008 GAGATAAAAAACACACTAAAAAACCAAGGAAAAAAGCTTTAAAGTTTCAACAT 114059
```

Search completed: January 9, 2003, 12:09:21  
Job time : 6395 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2003, 16:59:59 ; Search time 41 Seconds  
(without alignments)  
171.297 Million cell updates/sec

Title: US-09-438-185a-1047

Perfect score: 1889

Sequence: 1 VHYCERTLPKYLKALKL.....ESIPLYNQEKYLSGFVLCQ 362

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 118974 seqs, 19401057 residues

Total number of hits satisfying chosen parameters: 118974

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Published Applications AA:\*\*

1: /cgn2\_6/prodata/1/pubaa/PCT\_NEW\_PUB.pdb:\*\*

2: /cgn2\_6/prodata/1/pubaa/PCT\_NEW\_PUB.pdb:\*\*

3: /cgn2\_6/prodata/1/pubaa/US06\_NEW\_PUB.pdb:\*\*

4: /cgn2\_6/prodata/1/pubaa/US06\_PUBCOMB.pdb:\*\*

5: /cgn2\_6/prodata/1/pubaa/US07\_NEW\_PUB.pdb:\*\*

6: /cgn2\_6/prodata/1/pubaa/US07\_PUBCOMB.pdb:\*\*

7: /cgn2\_6/prodata/1/pubaa/PCTUS\_PUBCOMB.pdb:\*\*

8: /cgn2\_6/prodata/1/pubaa/US08\_PUBCOMB.pdb:\*\*

9: /cgn2\_6/prodata/1/pubaa/US09\_NEW\_PUB.pdb:\*\*

10: /cgn2\_6/prodata/1/pubaa/US10\_NEW\_PUB.pdb:\*\*

11: /cgn2\_6/prodata/1/pubaa/US10\_PUBCOMB.pdb:\*\*

12: /cgn2\_6/prodata/1/pubaa/US60\_NEW\_PUB.pdb:\*\*

13: /cgn2\_6/prodata/1/pubaa/US60\_PUBCOMB.pdb:\*\*

14: /cgn2\_6/prodata/1/pubaa/US60\_PUBCOMB.pdb:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	216.5	11.5	484	9 US-10-154-674-8	Sequence 8, Appli
2	216.5	11.5	485	9 US-10-154-674-6	Sequence 6, Appli
3	216.5	11.5	486	9 US-10-154-674-4	Sequence 4, Appli
4	216.5	11.5	490	9 US-10-154-674-2	Sequence 2, Appli
5	206.5	10.9	532	10 US-09-205-658-313	Sequence 313, App
6	95	5.0	1427	10 US-09-801-368-354	Sequence 354, App
7	92	4.9	427	10 US-09-817-774-23	Sequence 23, Appl
8	92	4.9	4563	9 US-09-870-759-128	Sequence 128, App
9	88.5	4.7	511	10 US-09-746-491-37	Sequence 37, Appl
10	88.5	4.7	550	10 US-09-746-491-40	Sequence 40, Appl
11	87.5	4.6	541	10 US-09-746-491-39	Sequence 39, Appl
12	87.5	4.6	541	12 US-10-157-447-2	Sequence 2, Appli
13	87.5	4.6	978	10 US-09-864-761-43207	Sequence 43207, A
14	84	4.4	456	10 US-09-741-669-385	Sequence 385, App
15	83.5	4.4	435	10 US-08-801-574-23	Sequence 22, Appl
16	83.5	4.4	435	10 US-08-801-574-23	Sequence 22, Appl
17	82.5	4.4	995	10 US-09-864-761-43017	Sequence 43017, A
18	82.5	4.4	1202	10 US-09-864-761-43061	Sequence 43061, A
19	82	4.3	909	10 US-09-815-242-12990	Sequence 12990, A

20	81.5	4.3	818	10 US-09-833-790-366	Sequence 366, App
21	81.5	4.3	1216	10 US-09-815-242-13312	Sequence 13312, A
22	81.5	4.3	1216	10 US-09-815-242-13698	Sequence 13698, A
23	81.5	4.3	3056	8 US-08-984-090-2	Sequence 2, Appli
24	81	4.3	323	10 US-09-801-368-352	Sequence 352, App
25	81	4.3	474	10 US-09-765-272-60	Sequence 60, Appl
26	81	4.3	823	9 US-09-895-913A-244	Sequence 244, App
27	81	4.3	1993	9 US-10-098-979-2	Sequence 2, Appli
28	80.5	4.3	197	10 US-09-764-853-836	Sequence 836, App
29	80.5	4.3	736	9 US-09-978-295A-526	Sequence 526, App
30	80.5	4.3	736	9 US-09-978-697-526	Sequence 526, App
31	80.5	4.3	736	9 US-09-978-192A-526	Sequence 526, App
32	80.5	4.3	736	9 US-09-999-832A-526	Sequence 526, App
33	80.5	4.3	736	9 US-09-978-189-526	Sequence 526, App
34	80.5	4.3	736	12 US-10-052-586-430	Sequence 420, App
35	80.5	4.3	765	10 US-09-888-615-80	Sequence 80, Appl
36	80	4.2	560	10 US-09-815-242-13304	Sequence 13304, A
37	80	4.2	897	10 US-09-815-242-11609	Sequence 11609, A
38	79	4.2	645	10 US-09-731-231A-2	Sequence 2, Appli
39	79	4.2	867	9 US-09-982-598-84	Sequence 84, Appl
40	79	4.2	867	9 US-09-989-293A-84	Sequence 84, Appl
41	79	4.2	867	9 US-09-989-735-84	Sequence 84, Appl
42	79	4.2	867	9 US-09-990-444-84	Sequence 84, Appl
43	79	4.2	867	9 US-09-989-730-84	Sequence 84, Appl
44	79	4.2	867	9 US-09-990-436-84	Sequence 84, Appl
45	79	4.2	867	9 US-09-991-181-84	Sequence 84, Appl
46	79	4.2	867	9 US-09-993-687-84	Sequence 84, Appl
47	79	4.2	867	9 US-09-989-734-84	Sequence 84, Appl
48	79	4.2	867	9 US-10-028-072-20	Sequence 20, Appl
49	79	4.2	867	10 US-09-989-723-84	Sequence 84, Appl
50	79	4.2	867	10 US-09-989-723-84	Sequence 84, Appl
51	79	4.2	867	10 US-09-989-727-84	Sequence 84, Appl
52	79	4.2	867	10 US-09-989-727-84	Sequence 84, Appl
53	79	4.2	867	10 US-09-989-731-84	Sequence 84, Appl
54	79	4.2	867	10 US-09-989-732-84	Sequence 84, Appl
55	79	4.2	867	10 US-09-991-073-84	Sequence 84, Appl
56	79	4.2	867	10 US-09-990-442-84	Sequence 84, Appl
57	79	4.2	867	10 US-09-991-163-84	Sequence 84, Appl
58	79	4.2	867	10 US-09-991-604-84	Sequence 84, Appl
59	79	4.2	867	10 US-09-990-456-84	Sequence 84, Appl
60	79	4.2	867	10 US-09-989-721-84	Sequence 84, Appl
61	78.5	4.2	410	10 US-09-881-752A-140	Sequence 140, App
62	78	4.1	185	9 US-10-007-270-15	Sequence 15, Appl
63	78	4.1	328	10 US-09-942-446-4	Sequence 4, Appli
64	78	4.1	328	10 US-09-795-693-35	Sequence 35, Appl
65	78	4.1	328	10 US-09-938-970-4	Sequence 4, Appli
66	77.5	4.1	506	10 US-09-815-242-5045	Sequence 5045, Ap
67	77.5	4.1	506	10 US-09-815-242-11757	Sequence 11757, A
68	77.5	4.1	790	10 US-09-801-368-184	Sequence 184, App
69	77	4.1	639	10 US-09-815-242-5390	Sequence 5390, Ap
70	77	4.1	646	10 US-09-815-242-12304	Sequence 12304, A
71	77	4.1	749	9 US-10-045-792-7	Sequence 7, Appli
72	77	4.1	749	10 US-09-833-205-6	Sequence 6, Appli
73	77	4.1	1663	10 US-09-875-519A-22	Sequence 22, Appl
74	77	4.1	1900	10 US-09-801-368-390	Sequence 390, App
75	76.5	4.0	311	9 US-10-027-805-41	Sequence 41, Appl
76	76.5	4.0	830	9 US-09-903-410-41	Sequence 41, Appl
77	76.5	4.0	830	9 US-10-071-505-3	Sequence 3, Appli
78	75.5	4.0	745	10 US-09-815-242-12209	Sequence 12209, A
79	75.5	4.0	833	9 US-09-847-010-4	Sequence 71, Appl
80	75	4.0	342	9 US-09-841-693-11	Sequence 11, Appl
81	75	4.0	396	10 US-09-841-693-9	Sequence 9, Appli
82	75	4.0	407	10 US-10-007-270-28	Sequence 28, Appl
83	75	4.0	777	9 US-10-007-270-2	Sequence 2, Appli
84	75	4.0	797	9 US-09-801-368-304	Sequence 304, App
85	75	4.0	1155	10 US-09-801-368-6	Sequence 6, Appli
86	75	4.0	1257	10 US-09-738-363-6	Sequence 118, App
87	75	4.0	1659	10 US-09-801-368B-118	Sequence 2, Appli
88	74.5	3.9	396	10 US-09-801-876B-2	Sequence 63, App
89	74.5	3.9	833	9 US-10-033-297-69	Sequence 22, Appl
90	74.5	3.9	1421	9 US-10-108-605-323	Sequence 56, Appl
91	74.5	3.9	2167	10 US-09-801-368-56	Sequence 33879, A
92	74	3.9	263	10 US-09-864-761-33879	

93	74	3.9	579	9	US-09-738-682-6373	Sequence 6373, App
94	74	3.9	579	10	US-09-965-825-2	Sequence 2, App1
95	74	3.9	579	10	US-09-965-825-5	Sequence 5, App1
96	73.5	3.9	229	10	US-09-828-664-113	Sequence 113, App
97	73.5	3.9	272	10	US-09-815-242-1040	Sequence 1040, App
98	73.5	3.9	311	10	US-09-866-065-145	Sequence 145, App
99	73.5	3.9	310	10	US-09-923-246-89	Sequence 89, App
100	73.5	3.9	547	10	US-09-961-5274-11	Sequence 11, App1
101	73.5	3.9	560	10	US-10-071-505-2	Sequence 2, App1
102	73.5	3.9	832	9	US-10-033-297-4	Sequence 4, App1
103	73.5	3.9	832	9	US-09-872-834-2	Sequence 2, App1
104	73.5	3.9	832	9	US-10-071-505-1	Sequence 1, App1
105	73.5	3.9	832	9	US-10-081-806-4	Sequence 4, App1
106	73.5	3.9	163	9	US-10-033-297-66	Sequence 66, App1
107	73	3.9	187	9	US-10-001-857-192	Sequence 192, App
108	73	3.9	585	10	US-09-815-242-11753	Sequence 11753, App
109	73	3.9	623	10	US-09-764-870-776	Sequence 276, App
110	73	3.9	649	9	US-10-063-547-132	Sequence 132, App
111	73	3.9	649	9	US-10-004-551-6	Sequence 6, App1
112	73	3.9	649	12	US-10-006-867-132	Sequence 132, App
113	73	3.9	649	12	US-10-005-868-184	Sequence 384, App
114	72.5	3.8	256	10	US-09-815-242-1336	Sequence 1336, App
115	72.5	3.8	394	10	US-09-923-302-122	Sequence 122, App
116	72.5	3.8	560	10	US-09-801-574-53	Sequence 53, App
117	72.5	3.8	569	12	US-10-078-929-116	Sequence 116, App
118	72.5	3.8	639	10	US-09-792-451-2	Sequence 2, App1
119	72.5	3.8	639	10	US-09-792-451-2	Sequence 2, App1
120	72.5	3.8	639	12	US-10-052-586-364	Sequence 364, App1
121	72.5	3.8	708	10	US-09-735-705-369	Sequence 369, App
122	72.5	3.8	708	10	US-09-850-7164-369	Sequence 369, App
123	72.5	3.8	878	10	US-09-897-778-369	Sequence 369, App
124	72.5	3.8	878	10	US-09-950-370-122	Sequence 22, App1
125	72.5	3.8	1167	10	US-09-815-242-11522	Sequence 11522, App
126	72	3.8	1329	10	US-09-775-879-17	Sequence 17, App1
127	72	3.8	1648	10	US-09-738-626-4195	Sequence 4195, App
128	71.5	3.8	304	9	US-09-738-626-4195	Sequence 18, App1
129	71.5	3.8	457	9	US-10-136-517-18	Sequence 136, App1
130	71	3.8	176	10	US-09-864-751-33919	Sequence 33919, App
131	71	3.8	309	10	US-09-868-055-489	Sequence 489, App
132	71	3.8	677	10	US-09-743-763-168	Sequence 168, App
133	71	3.8	759	10	US-09-764-864-823	Sequence 823, App
134	70.5	3.7	268	10	US-09-841-112-832	Sequence 832, App
135	70.5	3.7	321	10	US-09-886-055-117	Sequence 117, App1
136	70.5	3.7	351	10	US-09-933-943-26	Sequence 26, App
137	70.5	3.7	384	10	US-09-729-674-174	Sequence 174, App1
138	70.5	3.7	455	10	US-09-971-956-30	Sequence 30, App
139	70.5	3.7	524	10	US-09-925-300-150	Sequence 150, App
140	70.5	3.7	585	10	US-09-815-242-5426	Sequence 1538, App
141	70.5	3.7	585	10	US-09-815-242-12620	Sequence 12620, App
142	70.5	3.7	631	9	US-10-108-605-169	Sequence 169, App
143	70.5	3.7	737	10	US-09-749-601A-9	Sequence 9, App1
144	70.5	3.7	884	9	US-09-843-676-55	Sequence 55, App1
145	70.5	3.7	884	9	US-09-766-253-55	Sequence 55, App1
146	70.5	3.7	1042	10	US-09-815-242-13345	Sequence 13345,

```

Sequence 6373, Ap
Sequence 2, Appl
Sequence 5, Appl
Sequence 113, App
Sequence 10490, A
Sequence 145, App
Sequence 89, App
Sequence 11, Appl
Sequence 2, Appl
Sequence 4, Appl
Sequence 1, Appl
Sequence 4, Appl
Sequence 66, Appl
Sequence 192, App
Sequence 11753, A
Sequence 276, App
Sequence 132, App
Sequence 6, Appl
Sequence 132, App
Sequence 384, App
Sequence 13326, A
Sequence 722, App
Sequence 53, Appl
Sequence 116, App
Sequence 2, Appl
Sequence 364, App
Sequence 369, App
Sequence 369, App
Sequence 369, App
Sequence 22, Appl
Sequence 11522, A
Sequence 17, Appl
Sequence 4, Appl
Sequence 4195, Ap
Sequence 18, Appl
Sequence 33919, A
Sequence 499, App
Sequence 168, App
Sequence 823, App
Sequence 433, App
Sequence 117, App
Sequence 26, Appl
Sequence 174, App
Sequence 30, Appl
Sequence 1538, App
Sequence 5426, Ap
Sequence 12620, A
Sequence 169, App
Sequence 9, Appl
Sequence 55, Appl
Sequence 55, Appl
Sequence 13745, A
Sequence 249, App
Sequence 39, App
Sequence 39, Appl
Sequence 73, Appl

CURRENT APPLICATION NUMBER: US/10/154,674
CURRENT FILING DATE: 2002-05-23
PRIOR APPLICATION NUMBER: US 60/294,076
PRIOR FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8
LENGTH: 464
TYPE: PRT
ORGANISM: homo sapiens
US-10-154-674-8

Query Match          11.5%; Score 216.5; DB 9; Length 464;
Best Local Similarity 26.2%; Pred. No. 5,4e-13;
Matches 67; Conservative 42; Mismatches 104; Indels 43; Gaps 4;

QY      117 LMKYCPRPFLDYLEAFGLSDPFDH-----QAVKPELETHTFYVSGFVAPHQ 168
DB      221 LSKYIPFHACRELYKNFKPLLTTCGYGRDNVPQLDEVSFLEKRGGFIVRVAAVLSPRD 280
QY      169 YLISLDQRFPPIASWMTLIDKNSLRPDLLIHLGHVHWMLHPSPSEFFINMRRLTKV 228
DB      261 FLAGLAIVRHCTQYIRHGSDPLYTPPEPTCHELGHVPLDDPFAPQSCEIG----- 334
QY      229 IEKVQALPSKKORIQTGQSNLIAIVRCFWPTVESGLIEHBKRAYAGVALISSPOELGHA 288
DB      335 ---LASTIGASDEDVKLAT-----CYFTIERGLCKEGOLRAVYAGALLSTIGELKHA 364
QY      289 FIDNRVLPLELDQILRIPLFNFTSPQETLFSIRHPDEIVE-----L 329
DB      365 LSDAACVKAEPDPKTTOECILTTFOEAYFVSESFEBAKEKMDPAKSITRPPSVENPY 444
QY      330 TSKLEMMLDQGLLES1 345
DB      445 TGSIILLIKDRSIEVN 460

RESULT 2
US-10-154-674-6
Sequence 6, Application US/10154674
Publication No. US20020192694A1
GENERAL INFORMATION:
APPLICANT: yu, Xuanchuan
APPLICANT: Miranda, Maricar
APPLICANT: Hu, Yi
TITLE OF INVENTION: No. US20020192694A1el Human Hydroxyjaees and Polynucleotides
FILE REFERENCE: LEX-0352-USA
CURRENT APPLICATION NUMBER: US/10/154,674
CURRENT FILING DATE: 2002-05-23
PRIOR APPLICATION NUMBER: US 60/294,076
PRIOR FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 465
TYPE: PRT
ORGANISM: homo sapiens
US-10-154-674-6
```

```

RESULT 1
US-10-154-674-8
; Sequence 8, Application US/10154674
; Publication No. US20020192694A1
; GENERAL INFORMATION:
; APPLICANT: Yu, Xuanchuan
; APPLICANT: Miranda, Maxicar
; APPLICANT: Hu, Yi
; TITLE OF INVENTION: NO. US20020192694A1 Human Hydroxylases and Polynucleotides Encoded by the Same Gene
; FILE REFERENCE: LEX: 0352-USA

```

```

; CURRENT APPLICATION NUMBER: US/10/154,674
; CURRENT FILING DATE: 2002-05-23
; PRIOR APPLICATION NUMBER: US 60/294,076
; PRIOR FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 484
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-154-674-8

Query Match 11.5%; Score 216.5; DB 9; Length 484;
Best Local Similarity 26.2%; Pred. No. 5,4e-13;
Matches 67; Conservative 42; Mismatches 104; Indels 43; Gaps 4;

QY 117 LMSYCRFLDYLEAFGLSDFLDH-----QAVIKFPELETHPSYVSGFVAHQ 168
DB 221 LSKLYPTHACRELYKNFLLTKYCCGYEDBNVPQLEBVSNFLERSGFTVRVAAVYLSPRD 280
QY 169 YLSLQDRYPPPLASVWRITLIDKDNFSLRPDLIHLLGHVWMLHPSEFSEFFINMGRLLTKV 228
DB 281 FLAGLAVRVHCTQYIRHGSDDLTPPEBDCHELLGHVPLADPKFAFQSOEID----- 334
QY 229 IEKQALPSKKORIQTLOSNIILAVRCFMTFVESGLIENHBEKRAKYGAVLLSPQELGHA 288
DB 335 ---LASIGASDEDEVKLAT-----CYFFTIEFGLCXQGOGLRAYGAGLLSIGELKHA 384
QY 289 FIDNVRVLPLELDQIIRLPFNTSTPOEFLFSIHNPDELYE-----L 329
DB 385 LSDKACVKAPADPRKTCLOECCLITTFQEAIVFVSSFEAKEKKRDPKASITRPFVSYPNY 444
QY 330 TSKLEWMLDQGLLESI 345
DB 445 TQSEILKQTRSIENV 460

RESULT 2
US-10-154-674-6
; Sequence 6, Application US/10154674
; Publication No. US20020192694A1
; GENERAL INFORMATION:
; APPLICANT: Yu, Xuanchuan
; APPLICANT: Miyazawa, Maricar
; APPLICANT: Hu, Yi
; TITLE OF INVENTION: No. US20020192694A1 Human Hydroxylases and Polynucleotides
; FILE REFERENCE: LEX-0352-USA
; CURRENT APPLICATION NUMBER: US/10/154,674
; CURRENT FILING DATE: 2002-05-23
; PRIOR APPLICATION NUMBER: US 60/294,076
; PRIOR FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 485
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-154-674-6

Query Match 11.5%; Score 216.5; DB 9; Length 485;
Best Local Similarity 26.2%; Pred. No. 5,4e-13;
Matches 67; Conservative 42; Mismatches 104; Indels 43; Gaps 4;

QY 117 LMSYCRFLDYLEAFGLSDFLDH-----QAVIKFPELETHPSYVSGFVAHQ 168
DB 222 LSKLYPTHACRELYKNFLLTKYCCGYEDBNVPQLEBVSNFLERSGFTVRVAAVYLSPRD 281
QY 169 YLSLQDRYPPPLASVWRITLIDKDNFSLRPDLIHLLGHVWMLHPSEFSEFFINMGRLLTKV 228
DB 282 FLAGLAVRVHCTQYIRHGSDDLTPPEBDCHELLGHVPLADPKFAFQSOEID----- 335
QY 229 IEKQALPSKKORIQTLOSNIILAVRCFMTFVESGLIENHBEKRAKYGAVLLSPQELGHA 288

```

```
Db 336 ---LASIGASDEVDOKLAT-----CYFTTIEFGLCKQEGQRAYGAGLLSSIGELKHA 385
Qy 289 FIDNVRVLPLELDQIIRLPFTNSTPQETLFSIRHDELVE-----L 329
Db 386 LSKACVAKAFDPKTTCLQECCLITTFQEAIFYSESFEAEKMRDFAKSIITRPPSVYFNPY 445
Qy 330 TSLEWMLDQGLLESI 345
Db 446 TQSIIEILKDRSTIENV 461

RESULT 3
US-10-154-674-4
; Sequence 4, Application US/10154674
; Publication No. US20020192694A1
; GENERAL INFORMATION:
; APPLICANT: Yu, Xuanchuan
; APPLICANT: Miranda, Maricar
; APPLICANT: Hu, Yi
; TITLE OF INVENTION: No. US20020192694A1el Human Hydroxylases and Polynucleotides Encc
; FILE REFERENCE: LEX-0352-USA
; CURRENT APPLICATION NUMBER: US/10/154,674
; PRIOR FILING DATE: 2002-05-23
; PRIOR APPLICATION NUMBER: US 60/294,076
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 486
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-154-674-4

Query Match 11.5%; Score 216.5; DB 9; Length 486;
Best Local Similarity 26.2%; Pred. No. 5.4e-13;
Matches 67; Conservative 42; Mismatches 104; Indels 43; Gaps 4;

Qy 117 LKWSYCPFRFDLYLEAFGLSDFLDH-----QAVIKFFFELETHFSYYPVSGFVAPHQ 168
Db 223 LSKLYPTHACREYLNKFPLLTKYCGYREDNVLPQEDVSMFLKERSGFTVRPVAGYLSPRD 282
Qy 169 YLSLQDRYPPPIASVMRTLDKDNFSLTPDLIHDLGHVPHLLHPSSEFFINMGRFLTQV 228
Db 283 FLAGLAYRVFCHCTQYIRHGSPLYTPEDTCHELLGHVPLLDAPKPAQFSQEI----- 336
Qy 229 IEKQALPSKQRIOTLOSNIIVRCFWFTVESGLIENHEGRKAYGAVLISSPOELGHA 288
Db 337 ---LASIGASDEVDOKLAT-----CYFTTIEFGLCKQEGQRAYGAGLLSSIGELKHA 386
Qy 289 FIDNVRVLPLELDQIIRLPFTNSTPQETLFSIRHDELVE-----L 329
Db 387 LSKACVAKAFDPKTTCLQECCLITTFQEAIFYSESFEAEKMRDFAKSIITRPPSVYFNPY 446
Qy 330 TSLEWMLDQGLLESI 345
Db 447 TQSIIEILKDRSTIENV 462

RESULT 4
US-10-154-674-2
; Sequence 2, Application US/10154674
; Publication No. US20020192694A1
; GENERAL INFORMATION:
; APPLICANT: Yu, Xuanchuan
; APPLICANT: Miranda, Maricar
; APPLICANT: Hu, Yi
; TITLE OF INVENTION: No. US20020192694A1el Human Hydroxylases and Polynucleotides Encc
; FILE REFERENCE: LEX-0352-USA
; CURRENT APPLICATION NUMBER: US/10/154,674
; PRIOR FILING DATE: 2002-05-23
; PRIOR APPLICATION NUMBER: US 60/294,076
; NUMBER OF SEQ ID NOS: 8
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 490
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-154-674-2

Query Match 11.5%; Score 216.5; DB 9; Length 490;
Best Local Similarity 26.2%; Pred. No. 5.5e-13;
Matches 67; Conservative 42; Mismatches 104; Indels 43; Gaps 4;

Qy 117 LKWSYCPFRFDLYLEAFGLSDFLDH-----QAVIKFFFELETHFSYYPVSGFVAPHQ 168
Db 227 LSKLYPTHACREYLNKFPLLTKYCGYREDNVLPQEDVSMFLKERSGFTVRPVAGYLSPRD 286
Qy 169 YLSLQDRYPPPIASVMRTLDKDNFSLTPDLIHDLGHVPHLLHPSSEFFINMGRFLTQV 228
Db 287 FLAGLAYRVFCHCTQYIRHGSPLYTPEDTCHELLGHVPLLDAPKPAQFSQEI----- 340
Qy 229 IEKQALPSKQRIOTLOSNIIVRCFWFTVESGLIENHEGRKAYGAVLISSPOELGHA 288
Db 341 ---LASIGASDEVDOKLAT-----CYFTTIEFGLCKQEGQRAYGAGLLSSIGELKHA 390
Qy 289 FIDNVRVLPLELDQIIRLPFTNSTPQETLFSIRHDELVE-----L 329
Db 391 LSKACVAKAFDPKTTCLQECCLITTFQEAIFYSESFEAEKMRDFAKSIITRPPSVYFNPY 450
Qy 330 TSLEWMLDQGLLESI 345
Db 451 TQSIIEILKDRSTIENV 466

RESULT 5
US-09-205-658-313
; Sequence 313, Application US/09205658
; Patent No. US20010029617A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Ory, Scott
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; FILE REFERENCE: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; CURRENT APPLICATION NUMBER: US/09/205,658
; CURRENT FILING DATE: 1998-12-03
; EARLIER APPLICATION NUMBER: 08/857,076
; EARLIER FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: 08/888,534
; EARLIER FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: US98/10080
; EARLIER FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 328
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 313
; LENGTH: 532
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-205-658-313

Query Match 10.9%; Score 206.5; DB 10; Length 532;
Best Local Similarity 25.2%; Pred. No. 5.6e-12;
Matches 70; Conservative 47; Mismatches 116; Indels 45; Gaps 8;

Qy 104 RNLM---YRLSSRFLSKSYCRFFDLYLEAFGLSDFLDH-----QAVIKFFFELETHF 155
Db 251 RKTWGIIVRKUR---ELKHGKACKQFLDNFELLERHCGYSENNIPOLEDICKFLKAKTGF 307
Qy 156 SYYPVSGFVAPHQYLSLQDRYPPPIASVMRTLDKDNFSLTPDLIHDLGHVPHLLHPSFS 215
Db 308 RVRFVAGYLSARDFLAGLAYRVFCHCTQYIRHGSPLYTPEDTCHELLGHVPLLDAPKPA 367
Qy 216 EFFINMGRFLTQVIEKVOALPSKQRIOTLOSNIIVRCFWFTVESGLI----- 265
Db 368 QFSQEI-----LASIGASDEVDOKLAT-----YFSTIEFGLSDSDADSPVK 411
```

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 4

QY 266 ---ENHEGRKAYGAVLISSEPOELGHAFIDNRVLPLELDIOIRLPENTSTPOETLFSIRH 322  
DB 412 ENSGNHERRKAYGAGLISAGELGJAVGSGTIIIRPPDKVAVOEGCITTFQSAFYFRN 471  
QY 323 FDELVELTSLKEMMDQGLLESIPLYNOEKLSGFEVL 360  
DB 472 FEEAQO---KLRFMTNMKRPYFRN--PYTSVEVL 504

RESULT 6  
US-09-801-368-354  
Sequence 354, Application US/09801368  
Patent No. US20020128250A1  
GENERAL INFORMATION:  
APPLICANT: Busby, Robert  
APPLICANT: Call, Brian  
APPLICANT: Hecht, Peter  
APPLICANT: Holtzman, Doug  
APPLICANT: Madden, Kevin  
APPLICANT: Maxon, Mary  
APPLICANT: Milne, Todd  
APPLICANT: No. US20020128250A1man, Thea  
APPLICANT: Royer, John  
APPLICANT: Salama, Soile  
APPLICANT: Sherman, Amir  
APPLICANT: Silva, Jeff  
APPLICANT: Summers, Eric  
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
FILE REFERENCE: 109272.147  
CURRENT APPLICATION NUMBER: US/09/801,368  
PRIORITY FILING DATE: 2001-03-07  
PRIORITY FILING DATE: 2000-01-19  
PRIORITY FILING DATE: 1999-10-20  
PRIORITY FILING DATE: 1999-10-20  
NUMBER OF SEQ ID NOS: 440  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 354  
LENGTH: 1427  
TYPE: PRF  
ORGANISM: Saccharomyces cerevisiae  
US-09-801-368-354

Query Match 5.0%; Score 95; DB 10; Length 1427;  
Best Local Similarity 18.8%; Pred. No. 1.1; Indels 204; Gaps 25;  
Matches 98; Conservative 65; Mismatches 155;  
QY 12 YIIKIALKTRQSLSPFONGSGLORAVSTPYSTRIT-----LQENKREKQALAHKCI 65  
DB 214 YILE---KLIFDMTHYNDQQ--RTWKRQISTYFLKILGNCYSRLINX---IFHMLV 266  
QY 66 SILEFFKNLFF---VHLTSLSKNORECSTDMAVST-----PFF----- 102  
DB 267 EFINKEWNEFFELPLSLHILMIFWMDICQIDINAVPAATITSSQKEPFLVTKITDMLHK 326  
QY 103 ---NENLWYRLSSRFSL-----WKS 120  
DB 327 YIIVSSSKMINDENYIINDIKONKIKLITLISLILKIFQOGLSEVFIPTSNWEI 386  
QY 121 YCPRF---LDYLEAFGLSDPFDHOAVIKFPELEHNSYVPGFVAP 166  
DB 387 YKELLEIYVSNADTQNSNMKKKLLISYRNEBKNNSSIR-----NYMSSNN 437  
QY 167 HOYLSLQDRYFPPIASVM-----RTLDKN-----FSLTPDLIHLIGHVP 207  
DB 438 DFOUITYTCQKPLSCIOJLNCITQFTKLLD-DNPEFPMPTVVDONPLTMKIIQLIL 496  
QY 208 WLLAPS--FSEFFINNGRLFTVIEKVOALPSKQRIQTLQSNL--IAIVRCFWFTVSG 263  
DB 497 WSIHPSRQPDHESN---OLVAKLLL-----RINSTDDELDHEFOIEDAIVSLVFO- 544  
QY 264 LIENHEGRK-----AYGAV-----LISSP---QELGHAVID-- 291

DB 545 LAKFSAQKQVSVWMPSLYRLNLTITGIIKVPYIRKLISSGLLYLQDSNDKRVHQ 604  
QY 232 ---NVRVLPLELDIOIRLPENTSTPOETLFSIRHDELVELTSLK----- 333  
DB 605 LINDKISPLMKQYMMVLRNVAEDVKEIIFNDQVLEITTEQIKRIRIISNDITMLQLS 664  
QY 334 -----EMMLD---QGLLESIPLYNOEKYISGFEVLC 361  
DB 665 KTLPLSIKIMVAEWYLSHLCSGLISV---NRTVLKIKIFKIPC 703

RESULT 7  
US-09-817-774-23  
Sequence 23, Application US/09817774  
Patent No. US2002012011A1  
GENERAL INFORMATION:  
APPLICANT: CHOE, Sungghwa  
APPLICANT: FELDMANN A., Kenneth  
TITLE OF INVENTION: Dwf5 MUTANTS  
FILE REFERENCE: 2225-0020 / 91020.002  
CURRENT APPLICATION NUMBER: US/09/817,774  
PRIORITY FILING DATE: 2001-03-26  
PRIORITY FILING DATE: 2000-03-27  
NUMBER OF SEQ ID NOS: 45  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 23  
LENGTH: 427  
TYPE: PRF  
ORGANISM: Artificial Sequence  
OTHER INFORMATION: Description of Artificial Sequence: LBR-RAT  
US-09-817-774-23

Query Match 4.9%; Score 92; DB 10; Length 427;  
Best Local Similarity 20.4%; Pred. No. 0.41;  
Matches 87; Conservative 53; Mismatches 138; Indels 148; Gaps 19;  
QY 41 PYSYRIILOKENEKQAL-----ARHKCISILEFFKNLFLVHLTSLSKNQ 86  
DB 35 PACVFPILLQCAQKDPGLQFPPLPALRELMENAVCGVILWFLQALF-SLLPYGK-- 91  
QY 87 REGCSTDMAYVSPFRN-RULWYRLSSRFSIMKSCYCRFPDLYLEAFGLSDPFDHOAV 145  
DB 92 -----VVEGTPLDVGRKLYRL-----NGLYAPILISAAY 121  
QY 146 -----IKFEELTHESYVPSGVAPHOYISLQDRYFPPIASVMRTLDKNFSLTPDL 198  
DB 122 GTAVFMDLERYIYTHFLQPALAIV-----FSVLSYLYVARSLLKVPRELSPASSGNA 176  
QY 199 IHDLLGHVPLMLHPSFS---EFFINM--GRLEFTVIEKVOALPSKQRIQTLQSNL 247  
DB 177 YVDFP--IGRELNRIGAFDLKFCFLRGILGWVIVLWMLLEMVQERSAPSLAMTL 234  
QY 248 ---NLIAIVRCFWFTVSGILEN---HEG---RKAYGAV-----LISSP 283  
DB 235 VNSFQLLYVVDALWF--EALTLTMDI IHDGFWLAFGDVWVPFTYSLOAFYLVNHQ 292  
QY 284 ELGHAVIDNRVLP-----ELDQIRLPENTSTPOETLFS- 319  
DB 293 DLSWPLSVIILKLCQYVIFRANCASQKNAFKNPTDKLAHLHTIP--ISTKSLVSG 350  
QY 320 ---IRHDELVELTSLKEMMDQGLLESIPLYN-----QEKYLSG 356  
DB 351 WKGFRHFNVTIGDILMALANSFLPCGNHILPFIYIYTTALLIHRARDEHOCRRKIGLA 410  
QY 357 FEVLQ 362  
DB 411 WEKYCQ 416  
RESULT 8

US-09-870-759-128

Sequence 128, Application US/09870759

Patent No. US20020177551A1

GENERAL INFORMATION:

APPLICANT: TERMAN, David S

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE

FILE REFERENCE: 870759

CURRENT APPLICATION NUMBER: US/09/870,759

CURRENT FILING DATE: 2002-01-14

PRIOR APPLICATION NUMBER: US 60/208,128

PRIOR FILING DATE: 2000-05-30

NUMBER OF SEQ ID NOS: 166

SOFTWARE: Patent in version 3.1

SEQ ID NO 128

LENGTH: 4563

TYPE: PRT

ORGANISM: Homo sapiens

US-09-870-759-128

Query Match 4.9%; Score 92; DB 9; Length 4563;

Best Local Similarity 20.7%; Pred No. 10; Indels 68; Gaps 12;

Matches 61; Conservative 48; Mismatches 118; Indels 68; Gaps 12;

QY 3 YCERYLDPKYILKALKRQSLFLFQNSQSLQRAYSTPYSYR---IILOKENKEKQAL 59

DB 4211 YTREELCTMFIREVGTVLQSVKVNQSEILL-----FSIQDLVITLPE-----L 4257

QY 60 ARHKCISILEFFKNLLFVHLLSLSKNQREGCSTDMVAVSTPFFNRLW-----YLLSS 113

DB 4258 RKHKLDIVSMYREL---LKDLKSAQEVKFAQSLKTTVL-RNIQDLLQIFQIED 4312

QY 114 RFLSKWYCPFFLDYL--BAFGLSDPDLHQAVIKFPELETHFSYYPVSGFVAP----- 166

DB 4313 NIKQLKEMKFTYLYNIQDEINTFIYIPY--VFKLKLENLCLNKHFNFTQNELQEA 4370

QY 167 -----HQYLSLQDRYFPFIASVMRT-----LDKDNFSLTPDLI----- 199

DB 4371 SQELQIQHYIMALREYFDPDSIVGWTVKYEELEKIVSLIKNLLVALKDPHSEIYIVSAS 4430

QY 200 ---HDLGLHPWLPHLHPSSEFFINMGLFTKVIKQVQALPFSKKQRIQFLQSNLIA 251

DB 4431 NFTSQLSSQVEQFLHNIQEVLSILTPDCKGKKEIAELSAQAQEI--IKSQALIA 4483

RESULT 9

US-09-746-491-37

Sequence 37, Application US/09746491

Patent No. US20020137202A1

GENERAL INFORMATION:

APPLICANT: Burgess, Catherine E.

TITLE OF INVENTION: No. US20020137202A1el Proteins and Nucleic Acids Encoding Same

FILE REFERENCE: 15966-621

CURRENT APPLICATION NUMBER: US/09/746,491

CURRENT FILING DATE: 2000-12-20

PRIOR APPLICATION NUMBER: USN 60/171,329

PRIOR FILING DATE: 1999-12-21

NUMBER OF SEQ ID NOS: 72

SOFTWARE: Patent in Ver. 2.0

SEQ ID NO 37

LENGTH: 511

TYPE: PRT

ORGANISM: Homo sapiens

US-09-746-491-37

Query Match 4.7%; Score 88.5; DB 10; Length 511;

Best Local Similarity 21.9%; Pred. No. 1.1;

Matches 84; Conservative 59; Mismatches 109; Indels 131; Gaps 19;

QY 25 SLFQNSQSLQRAYSTPYSYRIILOKENKEKQALRKHCISILEFFKNLLFVHLLSLSK 84

DB 4 SKLINSQQLLYQEYSD-----VVLNKEIQSQORLE-----SL 35

QY 85 NQREGCSTD---MAVVSTPFFNRLNLYLLSSRFSLW-----KS 120

Db 36 SETPGSPQPRKALVSESYLQRL---SMASSGSLWQEIIPVVRNSTVLLSMTHEDQKL 92  
QY 121 YCPREFLDYLEAFGL--LSDFLDHOAVIKFPELETHFSYYPVSGFVAPHQYL-SLLQDRY 177  
Db 93 QEVKFEIIVSEASYLSRLNIAVDH-----FQLSTSLR---ATLSNQEHOMLFSRLQDVR 143  
QY 178 FPIASVMRTDK--DNFSLTDPDLHDLGH-----VPMLLHPSFSE---FFINMGR 223  
Db 144 DVSNATFLSDLEENFENNIFSPQVCDVVLNHADPFRKRVLPYVTVNQTQERTFQSLMNSNS 203  
QY 224 LFTKVIKQVQALP-----SKK-----ORIOTLQSNLIAIVRCFCWFVTSGLIENHEGR 271  
Db 204 NFREVLEKLESDPVCORLSLKSFLILPQIRITRLKLLQNLK-----RTQPGSSEAEAT 259  
QY 272 KAYGAVLISSPOELGHAFIDNVRLVPLEDQIIRLPNTSTPQETLFSIRHFDDELVELTS 331  
Db 260 KAH-----HA-----LEQLIR-----DCNNNVQSMRTEELIYLSQ 290  
QY 332 KLEWMLDQGLLESIPLYNQEKYL 354  
Db 291 KIEFE-----CKIFFPLISQSRWL 308

RESULT 10

US-09-746-491-40

Sequence 40, Application US/09746491

Patent No. US20020137202A1

GENERAL INFORMATION:

APPLICANT: Burgess, Catherine E.

TITLE OF INVENTION: No. US20020137202A1el Proteins and Nucleic Acids Encoding Same

FILE REFERENCE: 15966-621

CURRENT APPLICATION NUMBER: US/09/746,491

CURRENT FILING DATE: 2000-12-20

PRIOR APPLICATION NUMBER: USN 60/171,329

PRIOR FILING DATE: 1999-12-21

NUMBER OF SEQ ID NOS: 72

SOFTWARE: Patent in Ver. 2.0

SEQ ID NO 40

LENGTH: 519

TYPE: PRT

ORGANISM: Homo sapiens

US-09-746-491-40

Query Match 4.7%; Score 88.5; DB 10; Length 519;

Best Local Similarity 21.9%; Pred. No. 1.2;

Matches 84; Conservative 59; Mismatches 109; Indels 131; Gaps 19;

QY 25 SLFQNSQSLQRAYSTPYSYRIILOKENKEKQALRKHCISILEFFKNLLFVHLLSLSK 84

Db 9 SKLINSQQLLYQEYSD-----VVLNKEIQSQORLE-----SL 40

QY 85 NQREGCSTD---MAVVSTPFFNRLNLYLLSSRFSLW-----KS 120

Db 41 SETPGSPQPRKALVSESYLQRL---SMASSGSLWQEIIPVVRNSTVLLSMTHEDQKL 97

QY 121 YCPREFLDYLEAFGL--LSDFLDHOAVIKFPELETHFSYYPVSGFVAPHQYL-SLLQDRY 177

Db 98 QEVKFEIIVSEASYLSRLNIAVDH-----FQLSTSLR---ATLSNQEHOMLFSRLQDVR 148

QY 178 FPIASVMRTDK--DNFSLTDPDLHDLGH-----VPMLLHPSFSE---FFINMGR 223

Db 149 DVSNATFLSDLEENFENNIFSPQVCDVVLNHADPFRKRVLPYVTVNQTQERTFQSLMNSNS 208

QY 224 LFTKVIKQVQALP-----SKK-----ORIOTLQSNLIAIVRCFCWFVTSGLIENHEGR 271

Db 209 NFREVLEKLESDPVCORLSLKSFLILPQIRITRLKLLQNLK-----RTQPGSSEAEAT 264

QY 272 KAYGAVLISSPOELGHAFIDNVRLVPLEDQIIRLPNTSTPQETLFSIRHFDDELVELTS 331

Db 265 KAH-----HA-----LEQLIR-----DCNNNVQSMRTEELIYLSQ 295

QY 332 KLEWMLDQGLLESIPLYNQEKYL 354

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 6

Db 296 KIEFE-----CKIFPLISQSRWL 313

RESULT 11

US-09-746-491-39  
Sequence 39, Application US/09746491  
Patent No. US20020137202A1  
GENERAL INFORMATION:  
APPLICANT: Burgess, Catherine E.  
TITLE OF INVENTION: No. US20020137202A1 Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 15966-621  
CURRENT APPLICATION NUMBER: US/09/746,491  
PRIOR FILING DATE: 2000-12-20  
PRIOR APPLICATION NUMBER: USSN 60/171,329  
NUMBER OF SEQ ID NOS: 72  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 39  
LENGTH: 550  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-746-491-39

Query Match 4.7%; Score 88.5; DB 10; Length 550;  
Best Local Similarity 21.9%; Pred No. 1.3; Indels 131; Gaps 19;  
Matches 84; Conservative 59; Mismatches 109; Indels 131; Gaps 19;

Qy 25 SLFFQNSQIQRAVSTPYRYRIITQENKQKQALAHKCSILEFFKALLFVHLISLSK 84  
Db 40 SKLINSQQLLYOEYSD-----VINKELIQSQORLE-----SL 71  
Qy 85 NQREGCSTD---MAVSTPFPRNLMWRLSSRFSIM-----XS 120  
Db 72 SETPGSPSPQPRKALVSSSESYQRL---SWASSGSLMOELPVRNSTVLLSMTHEDQTL 128  
Qy 121 YCRFFFLDYLEAFGL--LSDPLDHOAVIKFELETHFSYVSGFVAHQYL--SLQDRY 177  
Db 129 QEVVFELIVSEASYLSRLNIAVDH-----FOLSTSLR---ATLSNQHOMFLPSRLQDVR 179  
Qy 178 FLIASVWRTLDK--DNFSLTPDLIHDLGH-----VPMILHPSFS---FTINMR 223  
Db 180 DVATFTLSDEENFENNIFSPQCVVNLNAPDRRVLYPLVINGTQETPQSLNKS 239  
Qy 224 LPTKVIKQVALP-----SKT-----QRTQSQNLIAVRCFWTVESGLIENHGR 271  
Db 240 NFRVLEKLSDPVQRLSKSFLILPQRTIRKLKLQNLK---RTVPSSEAEAT 295  
Qy 272 KAYGAVLISPOELGHAFTDNVRLPLELQOILRLPFTSTPQETLFSIRHDELVELTS 331  
Db 296 KAH-----HA-----LEQLIR-----DCNNNVSMRTEELIYSQ 326  
Qy 332 KLEWMLDGLLESIPLYNOEKYL 354  
Db 327 KIEFE-----CKIFPLISQSRWL 344

RESULT 12

US-09-981-421-4  
Sequence 4, Application US/09981421  
Patent No. US20020098185A1  
GENERAL INFORMATION:  
APPLICANT: Sims, John E.  
APPLICANT: Mohler, Kendall M.  
TITLE OF INVENTION: METHODS FOR TREATING IL-18 MEDIATED DISORDERS  
FILE REFERENCE: 3086-A  
CURRENT APPLICATION NUMBER: US/09/981,421  
PRIOR FILING DATE: 2001-10-17  
PRIOR APPLICATION NUMBER: US 60/241,408  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: Patentin version 3.1

SEQ ID NO 4  
LENGTH: 541  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-981-421-4

Query Match 4.6%; Score 87.5; DB 10; Length 541;  
Best Local Similarity 21.3%; Pred No. 1.5; Indels 95; Gaps 19;  
Matches 68; Conservative 50; Mismatches 106; Indels 95; Gaps 19;

Qy 9 DPXYT---LKIALKLRQSLSPFQNSQGLQRAVSTPYRYRIITQENKQKQALAHK 64  
Db 34 EFPYIKHSCSLAEHETTKSWYKSGQCEHVELNPFSSRIAL-----HDC 81  
Qy 65 ISILEFF-----KNLFLVHLISLKNQREGCSTDMAVVS-----TPFN-- 103  
Db 82 --VLEFWEVLENDTGSYFFQMKNYTQKMLNVRIRNKHSCPTERQVTSKIVEYKFFOIT 139  
Qy 104 -RNLWRLSSRFSIMKSCPRFFLDYLEAFGLSDPLDHOAVIKFELETHFSYVPSG 162  
Db 140 CENSYYQTLVNSTSLYKN--CKKLLLENK-----NPTIKNAEPEDQ--GYSCVH 187  
Qy 163 FVAPHQYSLIQDRYPFIASVWRTLDKDFSLTPDLIHDLGHVPMILHPSFSFELINMG 222  
Db 188 FL--HH-----NGKLFYITK-----TFNIT--IVEDRSNIYPVLLGPKLHVAVELQ 230  
Qy 223 RLFTKVIKQVALPSKQRTQTSNLIAVRCFW--FTVESGLIEN--HEGRKAYGAVLS 280  
Db 231 K--NRLNCSALINEEDVI-----YMMFEGNSDPNTHEREK---NRLW 270  
Qy 281 SPQELGHAFTDNVRLPLE 299  
Db 271 TPEGKMAH---SKVLRIE 285

RESULT 13

US-10-157-447-2  
Sequence 2, Application US/10157447  
Patent No. US20020143155A1  
GENERAL INFORMATION:  
APPLICANT: Parnef, Patricia et al.  
TITLE OF INVENTION: Receptor Designated 2p1  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
STREET: 51 University Street  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple 7.1  
SOFTWARE: Microsoft Word, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/157,447  
FILING DATE: 28-May-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/578,178  
FILING DATE: <Unknown>  
APPLICATION NUMBER: 09/110,618  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Anderson, Kathryn A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2619  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
TELEX: 756822  
INFORMATION FOR SEQ ID NO: 2:



```

SEQUENCE CHARACTERISTICS:
  LENGTH: 541 amino acids
  TYPE: amino acid
  TOPOLOGY: linear
  MOLECULE TYPE: protein
  SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-157-447-2

Query Match 4.6%; Score 87.5; DB 12; Length 541;
Best Local Similarity 21.3%; Pred. No. 1.5;
Matches 68; Conservative 50; Mismatches 106; Indels 95; Gaps 19;

QY 9 DPKYI-----LKALKRQSLFQNSQSLQRAYSTPYYYRIILOKENEKQALARKHC 64
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 34 EFPYKHCSCSLAEIETTTKSWYKSGSQEHVLPNRRSSRIAL-----HDC 81
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 65 ISILEFF-----KNLLFVHLISLKNQREGCSTDMAVVS-----TPFFN-- 103
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 82 --VLEFPVELNDTGSYFFQMKYNTQKWLIRNRKHSCTERQVTSKIVEVKFFQIT 139
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 104 -RLNRYLSSRFLSKSYCPREFLDYLAFLGLSDFLDHQAIVKFELETHFSYYPVSG 162
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 140 CENSYQTLVNSTLYKN-CKLLENK-----NTIKNAEFEDQ-GYYSVCH 187
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 163 FVAPHOYLSLQDRFFFIASVMRTLDKNFESTPDLIHDLGHVPMLLHPSFSEFINNG 222
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188 FL--HH-----NGKLFNITK-----TFNIT--IVEDRNIYVPLGKLNHVAVEIG 230
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 223 RLFTVKIEKVALSKQRIOTLOSNIIVRCFW-FTVESGLIEN-HEGRKAYGAVLIS 280
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 231 K--NVRNLCSALLNEEDVI-----YWNFGENGSDPNHIEEKE-----WRIM 270
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 281 SPQELGHAFIDNVRVLPLE 299
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 271 TPEGKWH-----SKVLRIE 285
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
US-09-864-761-43207
  Sequence 43207, Application US/09864761
  Patent No. US20020048763A1
  GENERAL INFORMATION:
  APPLICANT: Penn, Sharon G.
  APPLICANT: Rank, David R.
  APPLICANT: Hanzel, David K.
  APPLICANT: Chen, Wensheng
  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
  TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
  FILE REFERENCE: Aecmica-X-1
  CURRENT APPLICATION NUMBER: US/09/864, 761
  CURRENT FILING DATE: 2001-05-23
  PRIOR APPLICATION NUMBER: US 60/180,312
  PRIOR FILING DATE: 2000-02-04
  PRIOR APPLICATION NUMBER: US 60/207,456
  PRIOR FILING DATE: 2000-05-26
  PRIOR APPLICATION NUMBER: US 09/632,366
  PRIOR FILING DATE: 2000-08-03
  PRIOR APPLICATION NUMBER: GB 24263.6
  PRIOR FILING DATE: 2000-10-04
  PRIOR APPLICATION NUMBER: US 60/236,359
  PRIOR FILING DATE: 2000-09-27
  PRIOR APPLICATION NUMBER: PCT/US01/00666
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00667
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00664
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00669
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00665
  PRIOR FILING DATE: 2001-01-30
  PRIOR APPLICATION NUMBER: PCT/US01/00668
  PRIOR FILING DATE: 2001-01-30

SEQUENCE CHARACTERISTICS:
  LENGTH: 541 amino acids
  TYPE: amino acid
  TOPOLOGY: linear
  MOLECULE TYPE: protein
  SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-157-447-2

Query Match 4.4%; Score 84; DB 10; Length 978;
Best Local Similarity 17.9%; Pred. No. 7.4;
Matches 53; Conservative 42; Mismatches 77; Indels 124; Gaps 10;

QY 8 LDPYKILKIALKRLQSLFQNSQSLQRAYST-----PYSYRIILOKENEKQALA 60
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 639 VDQYKUKAIDPKORELSFIVNSSVFLBEVISSELLKILYAFSHNMLVTENPDVVKLKT 698
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 R-----HKCIS-----TLEFFKNLLFV 77
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 699 RIVITLYNSIVLETTSEILVADNFDKNLCTSERYKEMVKQIVNSYGVKLDQYKSLIQI 758
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 78 HLISLSKNQREGCSTDMAVSTPFNNRLWYLLSSRFLSKSYCPREFLDYLAFLGLS 137
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 759 HRVIQS-----DTICFORKIYLLLEIY-----DY-QVQSLVS 791
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 138 DFLDHOAVIKFELETHFSYYPVSGVAPHOYLSLQDRFFFIASVMRTLDKNFSLTP- 196
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 792 G-----ELESSYSYPOADNI-----IRNLNITIKSHALPPY 825
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 197 -----DLIHDLGHVPMLH-----PSFSEFFINMGRLFTKVIK 232
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 826 ITVLPHSLLEDMVYRLHGHVFPSTHTENELKPKFPDPEFVEASKITDEIKEI 881
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
US-09-741-669-385
  Sequence 385, Application US/09741669
  Patent No. US20020022718A1
  GENERAL INFORMATION:
  APPLICANT: Forsyth, R. Allyn
  APPLICANT: Ohlsen, Kari L.
  APPLICANT: Zyskind, Judith W.
  TITLE OF INVENTION: Genes identified as required for
  TITLE OF INVENTION: proliferation of E. coli
  FILE REFERENCE: ELITRA.009A
  CURRENT APPLICATION NUMBER: US/09/741,669
  CURRENT FILING DATE: 2000-12-19
  PRIOR APPLICATION NUMBER: US 60/173005
  PRIOR FILING DATE: 1999-12-23
  NUMBER OF SEQ ID NOS: 481
  SOFTWARE: FastSeq for Windows Version 4.0
  SEQ ID NO 385
  LENGTH: 450
  TYPE: PPT
  ORGANISM: Escherichia coli
US-09-741-669-385

```

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

```
Query Match 4.44; Score 83.5; DB 10; Length 450;
Best Local Similarity 21.74; Pred No. 2.9; 102; Indels 67; Gaps 12;
Matches 57; Conservative 37; Mismatches 102; Indels 67; Gaps 12;

QY 106 LWRLLSRFSLKSKYCRFLDYLFARGLSDPLDHOAVIKFELEHTFSTYPSGVA 165
DB 215 LFGITRGIMISL-----MLAAGV-----GVGMFWLAKRYGMNVSGDEA 257
QY 166 PFOYSLLODRYF--IASVWRTLDKONFSLTPDLIHDLGHVPMWLPSPSEFFINMG 222
DB 258 FYFPLVLTDRTPSPWENLALLONVDNIDFQGLAPIVRDYVFIPSWLMPGPRSMVNSA 317
QY 223 RLFT-KVLEKVALPSKQRIQTLQSNLIA-----YVACFWFVESGLI 265
DB 318 NYFTWEVLNNHSGLAIS---PTLIGSLVWVGALFIPGLAIVVGLIIMFDMLEYELG-- 371
QY 266 ENHEGRKAYGAVALISSPQELGHAF---IDNRYVLPLE-LDQIIRLPFNTSTPQETLFSI 320
DB 372 -NREPRRYKAAIL-----HSFCFGAIFPMVTLAREGLDSFV-----SRVVFPI 413
QY 321 RHEDVELVETSKLEWMLDO-GLL 342
DB 414 VVFGACIMIAKLIVWLFESAGLI 436

RESULT 16
US-09-801-574-22
; Sequence 22, Application US/09801574
; Patent No. US20020081592A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Peijiang Jeremy
; TITLE OF INVENTION: Reproduction-specific Genes
; FILE REFERENCE: 0399.2007-002
; CURRENT APPLICATION NUMBER: US/09/801,574
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: 60/187,518
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: 60/261,557
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 22
; LENGTH: 835
; TYPE: prt
; ORGANISM: Mus musculus
US-09-801-574-22

Query Match 4.44; Score 83.5; DB 10; Length 835;
Best Local Similarity 24.44; Pred. No. 6.7;
Matches 47; Conservative 26; Mismatches 47; Indels 73; Gaps 10;

QY 210 LHPSPSEFFINMRLFTYVIEKVALP--SKQRIQTLQSNLIAIVRCFWFVESGL--- 264
DB 155 LHPSSSTFLHNVGLLENQFIKRRFFSLAKNEK---QSNLKDSIRDF---EALNVVC 206
QY 265 IENHEGR-KAYGAVALISSPQELGHAFIDNRYVLPLELDQIIRLPFNTSTPQETLFSIRHF 323
DB 207 ISNEKGERNVAREVDISKP---GFGF-----PFETNYPEDSGVDVADL 246
QY 324 DELVE-----LTSKLEW-----MLDQGL-----LES 344
DB 247 NDITKTLSPVLLETHCHENGLENHEHMKYTYLLYPELKWGLNNGVNTCYINVVLSQSCS 306
QY 345 IPIYNQEKLSGF 357
DB 307 IPIPIINDLPNQG 319

RESULT 17
US-09-864-761-49017
; Sequence 49017, Application US/09864761
```

```
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmtca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 49017
; LENGTH: 995
; TYPE: prt
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007282.2
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.6
; OTHER INFORMATION: EST HUMAN HIT: A1023682.1, EVALU8 8.00e-72
; OTHER INFORMATION: SWISSPROT HIT: P30622, EVALU8 1.00e-06
US-09-864-761-49017

Query Match 4.44; Score 82.5; DB 10; Length 995;
Best Local Similarity 20.94; Pred. No. 11;
Matches 70; Conservative 45; Mismatches 117; Indels 103; Gaps 15;

QY 78 HILSLSNRREGGSTM-----AVSPPPNR-----NL-----MYRL 111
DB 110 HLISEVNSKSSGTVHYIMKQIFVAPISLEIEYKSPSETPMNLQGLPTPKKSL 169
QY 112 SSFSLKSKYCRFPV-----DYLFARGLSDPLDHOAVIKFPEL--ETHFSTYPSV 161
DB 170 SSHTLPHENADEIELPQRSATSOIIQAFPI--DTLLSGIILKVELDKENHKSSLLGT 227
QY 162 GFVAHQYSLLODRYFPIASVWRTLDKONFSLTPDLIHDLGHVPMWLPSPSEFFINM 221
```



SEQ ID NO 12990  
LENGTH: 909  
TYPE: FRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-12990

Query Match 4.3%, Score 82; DB 10; Length 909;  
Best Local Similarity 22.4%; Pred. No. 10;  
Matches 87; Conservative 66; Mismatches 144; Indels 92; Gaps 22;

QY 11 KYLLKIALKLRQSLSPFQNSQSLORAVSTP---YSYRI--ILQKE-----NKEQA 58  
DB 512 KFOYSITKQOLYRTDLLFQFNDISKQTYVENLLINAGYQINLOQSLTIIVINQSKI 571  
QY 59 LARHKCISILEFFKNLLFVHLISLS---KNQRE-GCSTDMNAVSTPFFNRNLMYRLSSR 114  
DB 572 ----KTIPQNHIDNTQOQHQAALSWVKNERQAGATTD-----TFPGINKW--LIPIG 619  
QY 115 FSLMKSCYCRFLDYLEAFGLSDPLDHA--VIKFE--LETHSYYPVSGFVAPHOY 169  
DB 620 TSPFK-----GILA--IDYQSQVINPYDASILESMLESLA-----VEN 658  
QY 170 LSLIDRFFPIASWRTLDKNF--SLTPDL---IHLLGHVPMILHPSSEFFINMRL 224  
DB 659 VILLKQTRRESMLQAEKOLTHSNFLRSISHDRTLTITMGNLILVSHKD----- 709  
QY 225 FTKVIERQVALP---SKQRIQTLQSNLIAVRCFWTVE-----SGLIEN---HE 269  
DB 710 -MSIIEKQLLVHSFOESQVLYLVNTILSTKLQSSNVQIKQPYLSELVEIDMILE 768  
QY 270 GRKAYGAVLISSPOELGAFIDNVAVLPELIDQIIRLPFTSTPOETLSIRHPD-ELVE 328  
DB 769 RRLKKRITVSSSVNLQFHLIDSKLLOALFNLENAVKTSTDTKINISIRVASYEOIE 828  
QY 329 LTSKLEMLDQULESIPLYNOEKYLSGF 357  
DB 829 FA-----VIDEG--FGISLEQKIFEPF 850

RESULT 20  
US-09-833-790-366  
Sequence 366 Application US/09833790  
Patent No. US20020068288A1  
GENERAL INFORMATION:  
APPLICANT: Lodee, Michael J.  
APPLICANT: Wang, Tongcong  
APPLICANT: Secret, Heather  
APPLICANT: Monamach, Radoon  
APPLICANT: Indrias, Carol Y.  
APPLICANT: Fan, Liqun  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
FILE REFERENCE: 210121.512  
CURRENT APPLICATION NUMBER: US/09/833.790  
CURRENT FILING DATE: 2001-04-11  
NUMBER OF SEQ ID NOS: 440  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 366  
LENGTH: 818  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-833-790-366

Query Match 4.3%, Score 81.5; DB 10; Length 818;  
Best Local Similarity 19.8%; Pred. No. 10;  
Matches 57; Conservative 39; Mismatches 93; Indels 99; Gaps 12;

QY 95 AAVSTP-----FNRNLWYRLISSRFSJWS--YCPREFLDYLEAFGLSDF 139  
DB 27 AFVTIMCCPSRSSMLTGKRYVNHNVYNNENCSPPSQAMHBEKRTFAVYLNNTGYRTAF 86  
QY 140 LDHQAIVKFELETHSYYPVSGFVAPHOYLSLDQDRFFPIASWMT--LDDKNFSLTPD 197

DB 87 FG-----KYLN-BYNSYIP-PGW---REWGLIKNSRFYNTVCGNGIKKHGPDYAKD 136  
QY 198 LIHDL-----GHVPMILHPSSEFFINMCR----- 223  
DB 137 YFTDLITNESINFKMSKMYPRHRYVMVISHAAPHPEDSAIQSKLYNMQHITPBY 196  
QY 224 -----LTKVIERQVALPESKQRIQTLQSNLIAVRCFWTVE 260  
DB 197 NYAPNDKHWIMQYTPMLPIHMEFTNILQ-----RRLOTLMSVDSVRLNMLV 248  
QY 261 ESSLIEHNEGRKAYGAVLISSPOELGAFIDNVAVLPELIDQIIRLPF 308  
DB 249 ETGLEN-----TYIITYADHGHYHIGQGLVKGSMFEDPD--IRVPF 289

RESULT 21  
US-09-815-242-13312  
Sequence 13312 Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: Elitza, Ollia  
CURRENT APPLICATION NUMBER: US/09/815.242  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13312  
LENGTH: 1216  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-815-242-13312

Query Match 4.3%, Score 81.5; DB 10; Length 1216;  
Best Local Similarity 21.6%; Pred. No. 17;  
Matches 69; Conservative 46; Mismatches 123; Indels 81; Gaps 15;

QY 7 TLDPKYLKIALKLRQSLSPFQ-----NSQSLQRAVSTPYSYRRILO--KENKEMO 57  
DB 556 TGEKRLVIRKILKLDHQBKVAKFKETALLTSRSRNDQILLALBEYGIPIVKTDOEONNYAQ 615  
QY 58 ALAHKRCISILEFFKNLL-----FVYL-----LSLK----- 84  
DB 616 SLEQVQWLDLRYNIHPLDYLVALMMSPMFGDEDELARLSLQKAEKDHENLYEKV 675  
QY 85 NQREGCSTDMNAVST-----PFRNRLWYRLISSRFSJ--WKSYPREFLDYLEAF 133  
DB 676 NAOQWASSQGLHTHTLAELKQFMDILASWRLVAKTHSLYDLIWKIYDRFYDYV--- 732  
QY 134 GLSDPLDHAQIVKFELETHSYYPVSGFVAPHOYLSLDQ--DRYFPIASWMTLDKD 190

Db	733	GALPNGPARQA--NLVALALRADQEFKSNFKLSRFTIMIDQVLEAQHDLASVAVAPPKD	790
Qy	191	NFSLTPDLIHDLIG-HVPMLLHPSSEFFINMGRFLFTKVIKQVALPSKKQRIQTLQSNL	249
Db	791	AVELM--TIHKSGLLEFPYV-----FILNMQDFNKQDSMSEVILSRQ----	838
Qy	250	IAIVRCFWFTVESGLIENH	268
Db	839	IA-----KMETGAVEDH	850
RESULT 22			
US	09-815-242-13698	Application US/09815242	
Sequence 13698, Patent No. US2002061369A1			
GENERAL INFORMATION:			
APPLICANT: Haselbeck, Robert			
APPLICANT: Ohlsen, Karl L.			
APPLICANT: Zykind, Judith W.			
APPLICANT: Wall, Daniel			
APPLICANT: Trawick, John D.			
APPLICANT: Carr, Grant J.			
APPLICANT: Yamamoto, Robert T.			
APPLICANT: Xu, H. Howard			
TITLE OF INVENTION: Identification of Essential Genes in			
PROKARYOTES			
FILE REFERENCE: ELITRA.011A			
CURRENT APPLICATION NUMBER: US/09/815,242			
CURRENT FILING DATE: 2001-03-21			
PRIOR APPLICATION NUMBER: 60/191,078			
PRIOR FILING DATE: 2000-03-21			
PRIOR APPLICATION NUMBER: 60/206,848			
PRIOR FILING DATE: 2000-05-23			
PRIOR APPLICATION NUMBER: 60/207,727			
PRIOR FILING DATE: 2000-05-26			
PRIOR APPLICATION NUMBER: 60/242,578			
PRIOR FILING DATE: 2000-10-23			
PRIOR APPLICATION NUMBER: 60/253,625			
PRIOR FILING DATE: 2000-11-27			
PRIOR APPLICATION NUMBER: 60/257,931			
PRIOR FILING DATE: 2000-12-22			
PRIOR APPLICATION NUMBER: 60/269,308			
PRIOR FILING DATE: 2001-02-16			
NUMBER OF SEQ ID NOS: 14110			
SOFTWARE: FastSeq for Windows Version 4.0			
SEQ ID NO 13698			
LENGTH: 1216			
TYPE: PRT			
ORGANISM: Streptococcus pneumoniae			
US-09-815-242-13698			
Query Match 4.3%; Score 81.5; DB 10; Length 1216;			
Best Local Similarity 21.6%; Pred. No. 17;			
Matches 69; Conservative 46; Mismatches 123; Indels 81; Gaps 15;			
Qy	7	TLDPKYILKALKRQSLFFQ-----NSQSLQRAYSTPYVYRIILO---KENKEKQ	57
Db	556	TEGRLVKEILKLHKEGVAKPEIATLTSRSRNDQILLALSEYGPVKTDGEGNNYLQ	615
Qy	58	ALARKHCISILEFPKNILFVHLLSLSKNOPEGCSTDMVSTPFNNRLMYRLSSRFSL	84
Db	616	SLEVOVMDLTVRIHNPLODYALVALMSPMFGFDEDLARLSLQAKEDKHENLYEKIV	675
Qy	85	NORECSTDMVSTPFNNRLMYRLSSRFSL-----WKSVCPRFPDLYLEAF	133
Db	676	NAQWASSQKGLIHATAEKLKQFMDILASRLYKATSHYDLIKYINDRYDYDV---	732
Qy	134	GLLSDFLDHQAVIKFPELETHFSYPVSGFVAPHQYLSLQ---DRYFIASVMRTLDK	190
Db	733	GALPNGPARQA--NLVALALRADQEFKSNFKLSRFTIMIDQVLEAQHDLASVAVAPPKD	790
Qy	191	NFSLTPDLIHDLIG-HVPMLLHPSSEFFINMGRFLFTKVIKQVALPSKKQRIQTLQSNL	249

```

Db 1587 PSLBLBEINH-FLSVSVYDALPLTRLEGLKOLRQLYHDKDQVMDIMRASQONPDQIMV 1645
Qy 267 -----NHEGRK-----AYGAVL 278
Db 1646 KLVVNLQLSKRWAINHTGEKYLEAVSSCL 1675

```

```

RESULT 24
US-09-801-368-352
; Sequence 352, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Call, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Sumners, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 352
; LENGTH: 323
; TYPE: PRF
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-352

```

```

Query Match 4.3%; Score 81; DB 10; Length 323;
Best Local Similarity 23.0%; Pred. No. 3.2; Mismatches 86; Gaps 12;
Matches 56; Conservative 31; Indels 86; Gaps 12;

```

```

Qy 122 CRRFLDYL-EARGLSDPL-DHOAV--IKFELTHRSYVSGFVAPRQYLSLQDR 176
Db 134 CPOYIRTLVSEARTLMPETIPDPPTKVEFEFFYLLEESY-----LIVHHPQSLKO-- 186
Qy 177 YPIASVVR-----TLDKNLSLTPDLIHDLGHVWMLHPSFSEFFINMGRLLPTKVIE 230
Db 187 ---IVQVLRKOPPOITLSSDDQNCWSLINDSYINDVHLLYPPH----- 227
Qy 231 KYQALPSKKORIQTLSNLIIVRCFMFTVESGLIENHGRKAYGAVLSSPOELGHAFI 290
Db 228 -----ILVACLFTITI-----SIHKKPRKSSSLASASE----- 256
Qy 291 DNVRVLPLELDQIIRLPNTSTPOTLFSIRHPELVELSKLEMMLDQGLLESIPLYNQ 350
Db 257 -----AIRDPKSSSGPVQIARNRMASSIVD-----LEBWDI-IGEQTLIVDH 299

```

```

Qy 351 -EKY 353
Db 300 WDKY 303

RESULT 25
US-09-765-272-60
; Sequence 60, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.

```

```

; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 474 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-09-765-272-60

```

```

Query Match 4.3%; Score 81; DB 10; Length 474;
Best Local Similarity 20.2%; Pred. No. 5.4; Mismatches 113; Indels 76; Gaps 10;
Matches 59; Conservative 44; Mismatches 113; Indels 76; Gaps 10;

```

```

Qy 135 LLSPLDHO--AVIKF--FELTHRSYVSGFVAPRQYLSL-----LDRYPPIASVVR 185
Db 53 LTVKSSHKILRYIKVGSQNSYKTLPSSTYIAPSTKSNPFTIKDEKLFELIQOE 112
Qy 186 TLDKONSLSLTPDLIHDLGHV-----PMLHPSFS--EFTINMG 222
Db 113 LTAKNLDSLQSGRDTANELRIIVSEKLSAFNNFNQETKCLTETSFSPVPFANQVG 172
Qy 223 RLFTKVIKQVAL--PSKKORIQTLSNLIIVRCFMFTVESGLIENHGRKAYGAVLI 279
Db 173 ERFANLSLDLDITYKKARSDRVKQASBLR-----RVENELQKNRRKLLKQEKELL 225
Qy 280 SSP-----QELGHAFIDNVRVLPLELDQIIRLPNTSTP----- 313
Db 226 ATDNAEERFROKGBLLTTFHQVPPNDQVYIINDYTNQPIALDKALTEPQNAQRFFKR 285
Qy 314 -QELTFIRHPELVELT-----SKLEMMLDQGLLESIPLYNQEKYLSGF 357
Db 286 YQKKEAVKLTDLIETKATILLYESVETVYNQAGLEIAIREBELIQGF 337

```

```

RESULT 26
US-09-895-913A-244
; Sequence 244, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleanchous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Oomen, Raymond P.

```

; TITLE OF INVENTION: Identification of Polynucleotides  
 ; TITLE OF INVENTION: Genomic No. US20020160456Ael Helicobacter Polypeptides in the  
 ; FILE REFERENCE: 06132/043002  
 ; CURRENT APPLICATION NUMBER: US/09/895,913A  
 ; PRIOR FILING DATE: 2001-06-29  
 ; PRIOR APPLICATION NUMBER: US 08/881,227  
 ; PRIOR FILING DATE: 1997-06-24  
 ; NUMBER OF SEQ ID NOS: 368  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 244  
 ; LENGTH: 823  
 ; TYPE: PRT  
 ; ORGANISM: Helicobacter pylori  
 ; US-09-895-913A-244

Query Match 4.3%; Score 81; DB 9; Length 823;  
 Best Local Similarity 20.7%; Pred. No. 11;  
 Matches 59; Conservative 44; Mismatches 116; Indels 66; Gaps 14;

QY 17 ALKRSQSLSPFQNSQSRAYSTPYVYRIILOKENKEKQALAHKCI-SILEF---F 71  
 DB 559 SRLRLHLAGIFLEKSLAN-FYAWMPKNNLTKEYAETRTKLEKGVGAILDFGELGF 617  
 QY 72 KLLFVHLLSLSKNORE-----GCSTMAVSTP--FENRLWYLLSSRSLWKSVCPRF 125  
 DB 618 KGVLVETTAIVTQKSEKVLARSPLNLSIKQPSYIFDKQLPY-----WVIYRNAF 668  
 QY 126 P--LDYLEAFGLLSDFLDHQ-----AVIKFPELETHFSYYPVSGFVAPHQYLSL 172  
 DB 669 FDKVFSHQFGLFVDFRDITNSVLKNGIRVKSNDEN-----GKLIENIDSY 722  
 QY 173 LQDYFPIASVWRTLDKDNFSLTPDIHLLGHVPMWLLHPSFSFFINMGLRTKVIK 232  
 DB 723 IQEVLSPFKIASFLDRDDVLTNPNTYK-----PRILKKG-KGYVWNGS--VAILIPKN 774  
 QY 233 QALPSKKQ-----RIQTLOSNIIVRCFWFTV 260  
 DB 775 PISLSKKQCDVISSVEFRDFYKIARNYQTRTLN-IDSMSCFWFI 818

RESULT 27  
 US-10-098-979-2  
 ; Sequence 2, Application US/10098979  
 ; Patent No. US2002017207A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wyriad Genetics, Incorporated  
 ; APPLICANT: Sugiyama, Janice  
 ; APPLICANT: Cimbora, Daniel  
 ; TITLE OF INVENTION: TSG101-INTERACTING PROTEINS AND USE THEREOF  
 ; FILE REFERENCE: 1907.06  
 ; CURRENT APPLICATION NUMBER: US/10/098,979  
 ; PRIOR FILING DATE: 2002-03-14  
 ; PRIOR APPLICATION NUMBER: US 60/276,259  
 ; PRIOR FILING DATE: 2001-03-14  
 ; PRIOR APPLICATION NUMBER: US 60/304,101  
 ; PRIOR FILING DATE: 2001-07-10  
 ; PRIOR APPLICATION NUMBER: to be assigned  
 ; PRIOR FILING DATE: 2001-10-22  
 ; PRIOR APPLICATION NUMBER: to be assigned  
 ; PRIOR FILING DATE: 2002-01-07  
 ; NUMBER OF SEQ ID NOS: 2  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 2  
 ; LENGTH: 1993  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-098-979-2

Query Match 4.3%; Score 81; DB 9; Length 1993;  
 Best Local Similarity 18.0%; Pred. No. 38;  
 Matches 70; Conservative 61; Mismatches 121; Indels 136; Gaps 19;

QY 30 NSQSLQRAYSTPYVYRIILOKENKEKQALAHK--CISILEFFKNLLFVHLLSLKNOR 87  
 DB 1114 NSEIQRK-----LNEQELQDIEKHSTGVASVNLCEVLL-----HDC 1153  
 QY 88 EGGSTMAVSTPFFNRLN-----WYRL-----LSSRFSIWKSYCPRF--FLDYLEA 132  
 DB 1154 DACATDAECDSIQOATRNLDRRWRNICAMSMERELKIEETWRLWOKFLDDYSRPFEDWLK 1213  
 QY 133 FGLLSDFLDHQAVI-----KF--PELETHFSYYPVSGFVAPHQYLSLQDYFPIA 181  
 DB 1214 SERTRAPPSSSGVIYTVAKEELKKEAFORQVHECL-----TOLELINKQYRRLA 1263  
 QY 182 SVMRTLDKQNFSLTPDLIHDLLGHVPMWLLHPSSEFFINMGLRTKVIKVOALPSKKOR 241  
 DB 1264 RENET---DSACSLKQWHE--GNORW-----DNLOKRVTSILRLKHFIFIGOREE 1308  
 QY 242 IQTLOSNIIVRCFWFT-----VESGLIENH---EGRK 272  
 DB 1309 FETARDSILV-----WLTMDQLTNIHFSECDVQAKIKQLKAFQOEISLHNKIEQII 1363  
 QY 273 AYGAVLISSPQELGHAFIDNVRVLELDQIIRLPNTSTPQETLFSI-RHFDDELVELTS 331  
 DB 1364 AQGEQLIEKSEPLDAAIIEE-----ELDELRY-----COEVPGRVERHYHKKLIRL-- 1409  
 QY 332 KLEWMLDQGLLESIPLYNQEKYLSGFV 359  
 DB 1410 -----PLPDDHDLSDREL 1423

RESULT 28  
 US-09-764-853-836  
 ; Sequence 836, Application US/09764853  
 ; Patent No. US20020090672A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: F0206  
 ; CURRENT APPLICATION NUMBER: US/09/764,853  
 ; CURRENT FILING DATE: 2001-01-17  
 ; Prior application data removed - consult PALM or file wrapper  
 ; NUMBER OF SEQ ID NOS: 939  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 836  
 ; LENGTH: 197  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-09-764-853-836

Query Match 4.3%; Score 80.5; DB 10; Length 197;  
 Best Local Similarity 23.8%; Pred. No. 1.8;  
 Matches 30; Conservative 22; Mismatches 41; Indels 33; Gaps 5;  
 QY 197 DLIHLLGHVPMWLLHP-----SFSFFINMGLRLE---TKVI-EKVQALPSKKQ 240  
 DB 10 DALYDMIGFPDFILEPKELDDVDVGVGEISDSFFQNMNLNLYNFSKVMADQLRKPFRDQ 69  
 QY 241 RIQTLOSNIIVRCFWFTVESGLI-----ENHEGRKAYGAVLISSPQELGHA 288  
 DB 70 WMTTQT-----VNAYVLPFKNEIVFPAGILOAPFYARNHPKALNFGGIGVWGHETHA 124  
 QY 289 FIDNVR 294  
 DB 125 FDDQGR 130

RESULT 29  
 US-09-978-295A-526  
 ; Sequence 526, Application US/09978295A  
 ; Patent No. US20020156006A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ashkenazi, Avi  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Botstein, David

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 14

APPLICANT: Desnoyers, Luc  
APPLICANT: Baton, Dan  
APPLICANT: Ferrara, Napoleon  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gertsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, J. Christopher  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth J.  
APPLICANT: Kijavlin, Ivar J.  
APPLICANT: Kuo, Sophia S.  
APPLICANT: Napier, Mary A.  
APPLICANT: Pan, James;  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Shelton, David L.  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: P2630P1C11  
CURRENT FILING DATE: 2001-10-15  
PRIOR APPLICATION NUMBER: US/09/978,295A  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/064249  
PRIOR FILING DATE: 1997-11-03  
PRIOR APPLICATION NUMBER: 60/065311  
PRIOR FILING DATE: 1997-11-13  
PRIOR APPLICATION NUMBER: 60/066364  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 60/077450  
PRIOR FILING DATE: 1998-03-10  
PRIOR APPLICATION NUMBER: 60/077632  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077641  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077649  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077791  
PRIOR FILING DATE: 1998-03-12  
PRIOR APPLICATION NUMBER: 60/078004  
PRIOR FILING DATE: 1998-03-13  
PRIOR APPLICATION NUMBER: 60/078086  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078936  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078910  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078939  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079294  
PRIOR FILING DATE: 1998-03-25  
PRIOR APPLICATION NUMBER: 60/079656  
PRIOR FILING DATE: 1998-03-26  
PRIOR APPLICATION NUMBER: 60/079664  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079689  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079663  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079728  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079786  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079920  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/079923  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/080105  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080107  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080165  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080194  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080327  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080328  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080333  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080334  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/081070  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081049  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081071  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081195  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081203  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081229  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081955  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081817  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081819  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081952  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081838  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082568  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082569  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082704  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082804  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082700  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082797  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082796  
PRIOR FILING DATE: 1998-04-23  
PRIOR APPLICATION NUMBER: 60/083336  
PRIOR FILING DATE: 1998-04-27  
PRIOR APPLICATION NUMBER: 60/083322  
PRIOR FILING DATE: 1998-04-28  
PRIOR APPLICATION NUMBER: 60/083392  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083495  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083496  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083499  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083545  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083554  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083558



PRIOR FILING DATE: 1998-04-29  
 PRIOR APPLICATION NUMBER: 60/083559  
 PRIOR FILING DATE: 1998-04-29  
 PRIOR APPLICATION NUMBER: 60/083500  
 PRIOR FILING DATE: 1998-04-29  
 PRIOR APPLICATION NUMBER: 60/083742  
 PRIOR FILING DATE: 1998-04-30  
 PRIOR APPLICATION NUMBER: 60/084366  
 PRIOR FILING DATE: 1998-05-05  
 PRIOR APPLICATION NUMBER: 60/084414  
 PRIOR FILING DATE: 1998-05-06  
 PRIOR APPLICATION NUMBER: 60/084441  
 PRIOR FILING DATE: 1998-05-06  
 PRIOR APPLICATION NUMBER: 60/084637  
 PRIOR FILING DATE: 1998-05-07  
 PRIOR APPLICATION NUMBER: 60/084639  
 PRIOR FILING DATE: 1998-05-07  
 PRIOR APPLICATION NUMBER: 60/084640  
 PRIOR FILING DATE: 1998-05-07  
 PRIOR APPLICATION NUMBER: 60/084598  
 PRIOR FILING DATE: 1998-05-07  
 PRIOR APPLICATION NUMBER: 60/084600  
 PRIOR FILING DATE: 1998-05-07  
 PRIOR APPLICATION NUMBER: 60/084627  
 PRIOR FILING DATE: 1998-05-07  
 PRIOR APPLICATION NUMBER: 60/084643  
 PRIOR FILING DATE: 1998-05-07  
 PRIOR APPLICATION NUMBER: 60/085339  
 PRIOR FILING DATE: 1998-05-13  
 PRIOR APPLICATION NUMBER: 60/085338  
 PRIOR FILING DATE: 1998-05-13  
 PRIOR APPLICATION NUMBER: 60/085323  
 PRIOR FILING DATE: 1998-05-13  
 PRIOR APPLICATION NUMBER: 60/085582  
 PRIOR FILING DATE: 1998-05-15  
 PRIOR APPLICATION NUMBER: 60/085700  
 PRIOR FILING DATE: 1998-05-15  
 PRIOR APPLICATION NUMBER: 60/085689  
 PRIOR FILING DATE: 1998-05-15  
 PRIOR APPLICATION NUMBER: 60/085579  
 PRIOR FILING DATE: 1998-05-15  
 PRIOR APPLICATION NUMBER: 60/085580  
 PRIOR FILING DATE: 1998-05-15  
 PRIOR APPLICATION NUMBER: 60/085573  
 PRIOR FILING DATE: 1998-05-15  
 PRIOR APPLICATION NUMBER: 60/085704  
 PRIOR FILING DATE: 1998-05-15  
 PRIOR APPLICATION NUMBER: 60/085697

Query Match 4.3%; Score 80.5; DB 9; Length 736;  
 Best Local Similarity 23.8%; Pred. No. 11;  
 Matches 30; Conservative 22; Mismatches 41; Indels 33; Gaps 5;  
 QY 197 DLHDLGLHVPMLHP-----SFSEFFINWGLF---TKVI-EKVQALPSKKQ 240  
 Db 464 DAYDMIGFPDILFKEKLDVYDGYISEDSFFQNMNLNFSKAWADQKPPSRDQ 523  
 QY 241 RIOTLQSNLIAIVCFWFTVESGLI-----ENHEGRKAYGAVLISSPQELGHA 288  
 Db 524 WSMTPQT-----VNAYILPTKNEIVFPAGILOAPFYARNHPKALNFGGIVGVWGHETHA 578  
 QY 289 FIDNVR 294  
 Db 579 FDOGR 584

RESULT 30  
 US-09-978-697-526  
 ; Sequence 526 Application US/09978697  
 ; Patent No. US20020169284A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ashkenazi, Avi  
 ; APPLICANT: Baker Kevin P.

APPLICANT: Botstein, David  
 APPLICANT: Desnoyers, Luc  
 APPLICANT: Eaton, Dan  
 APPLICANT: Ferrara, Napoleon  
 APPLICANT: Filvaroff, Ellen  
 APPLICANT: Fong, Sherman  
 APPLICANT: Gao, Wei-Qiang  
 APPLICANT: Gerber, Hanspeter  
 APPLICANT: Geritsen, Mary E.  
 APPLICANT: Goddard, Audrey  
 APPLICANT: Godowski, Paul J.  
 APPLICANT: Grimaldi, J. Christopher  
 APPLICANT: Gurney, Austin L.  
 APPLICANT: Hillan, Kenneth J.  
 APPLICANT: Kljavin, Ivar J.  
 APPLICANT: Kuo, Sophia S.  
 APPLICANT: Napier, Mary A.  
 APPLICANT: Pan, James  
 APPLICANT: Paoni, Nicholas F.  
 APPLICANT: Roy, Margaret Ann  
 APPLICANT: Shelton, David L.  
 APPLICANT: Stewart, Timothy A.  
 APPLICANT: Tumas, Daniel  
 APPLICANT: Williams, P. Mickey  
 APPLICANT: Wood, William I.  
 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
 FILE REFERENCE: P2630PIC27  
 CURRENT APPLICATION NUMBER: US/09/978,697  
 CURRENT FILING DATE: 2001-10-16  
 PRIOR FILING DATE: 2001-07-30  
 PRIOR APPLICATION NUMBER: 09/918585  
 PRIOR FILING DATE: 1997-10-17  
 PRIOR APPLICATION NUMBER: 60/062250  
 PRIOR FILING DATE: 1997-11-03  
 PRIOR APPLICATION NUMBER: 60/064249  
 PRIOR FILING DATE: 1997-11-13  
 PRIOR APPLICATION NUMBER: 60/065311  
 PRIOR FILING DATE: 1997-11-21  
 PRIOR APPLICATION NUMBER: 60/066364  
 PRIOR FILING DATE: 1998-03-10  
 PRIOR APPLICATION NUMBER: 60/077450  
 PRIOR FILING DATE: 1998-03-11  
 PRIOR APPLICATION NUMBER: 60/077632  
 PRIOR FILING DATE: 1998-03-11  
 PRIOR APPLICATION NUMBER: 60/077641  
 PRIOR FILING DATE: 1998-03-11  
 PRIOR APPLICATION NUMBER: 60/077649  
 PRIOR FILING DATE: 1998-03-11  
 PRIOR APPLICATION NUMBER: 60/077791  
 PRIOR FILING DATE: 1998-03-12  
 PRIOR APPLICATION NUMBER: 60/078004  
 PRIOR FILING DATE: 1998-03-13  
 PRIOR APPLICATION NUMBER: 60/078886  
 PRIOR FILING DATE: 1998-03-20  
 PRIOR APPLICATION NUMBER: 60/078936  
 PRIOR FILING DATE: 1998-03-20  
 PRIOR APPLICATION NUMBER: 60/078910  
 PRIOR FILING DATE: 1998-03-20  
 PRIOR APPLICATION NUMBER: 60/078939  
 PRIOR FILING DATE: 1998-03-20  
 PRIOR APPLICATION NUMBER: 60/079294  
 PRIOR FILING DATE: 1998-03-25  
 PRIOR APPLICATION NUMBER: 60/079656  
 PRIOR FILING DATE: 1998-03-26  
 PRIOR APPLICATION NUMBER: 60/079664  
 PRIOR FILING DATE: 1998-03-27  
 PRIOR APPLICATION NUMBER: 60/079689  
 PRIOR FILING DATE: 1998-03-27  
 PRIOR APPLICATION NUMBER: 60/079663  
 PRIOR FILING DATE: 1998-03-27  
 PRIOR APPLICATION NUMBER: 60/079728  
 PRIOR FILING DATE: 1998-03-27  
 PRIOR APPLICATION NUMBER: 60/079786

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 16

PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079920  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/079923  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/080105  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080107  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080165  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080194  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080327  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080328  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080333  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080334  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/081070  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081045  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081071  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081195  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081203  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081229  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081955  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081817  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081819  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081952  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081838  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/082566  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082569  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082704  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082804  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082700  
PRIOR FILING DATE: 1998-04-23  
PRIOR APPLICATION NUMBER: 60/082797  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082796  
PRIOR FILING DATE: 1998-04-23  
PRIOR APPLICATION NUMBER: 60/08336  
PRIOR FILING DATE: 1998-04-27  
PRIOR APPLICATION NUMBER: 60/083322  
PRIOR FILING DATE: 1998-04-28  
PRIOR APPLICATION NUMBER: 60/083392  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083495  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083496  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083499  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083545  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083554  
PRIOR FILING DATE: 1998-04-29

PRIOR APPLICATION NUMBER: 60/083558  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083559  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083500  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083742  
PRIOR FILING DATE: 1998-04-30  
PRIOR APPLICATION NUMBER: 60/08366  
PRIOR FILING DATE: 1998-05-05  
PRIOR APPLICATION NUMBER: 60/084414  
PRIOR FILING DATE: 1998-05-06  
PRIOR APPLICATION NUMBER: 60/084441  
PRIOR FILING DATE: 1998-05-06  
PRIOR APPLICATION NUMBER: 60/084637  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084639  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084640  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084598  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084600  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084627  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084643  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/085339  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085338  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085323  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085582  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085700  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085689  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085579  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085580  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085573  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085704  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085697

Query Match 4.3%; Score 80.5; DB 9; Length 736;  
Best Local Similarity 23.8%; Pred. No. 11;  
Matches 30; Conservative 22; Mismatches 41; Indels 33; Gaps 5;

QY 197 DLHDLGHVWMLHP-----SFEFPINNGRLF---TKVL-EKVQALFSSKQ 240  
DB 464 DAYDMIGFPDFLEKEHDDVDYGEIISBDSFFQNMNLNYSASAKVMDQLRKPPSRQ 523  
QY 241 RIOTLQSNLAIYRCFWFVBSGLI-----ENHGRKAYGAVLISSPOELGHA 288  
DB 524 WSMTPQT-----VNAYLPTKNEIYFPAGILQAPFYARNHPKALNFGIGVWGHETLTA 578  
QY 289 FIDNVR 294  
DB 579 FDDQGR 584

RESULT 31  
US-09-978-192a-526  
Sequence 526, Application US/09978192A  
Patent No. US2002017753A1  
GENERAL INFORMATION:  
APPLICANT: Ashkenazi, Avi

us-09-438-185a-1047.rapb

Fri Jan 10 10:56:11 2003

APPLICANT: Baker Kevin P.  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan  
APPLICANT: Ferrara, Napoleon  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, J. Christopher  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth J.  
APPLICANT: Klijavin, Ivar J.  
APPLICANT: Kuo, Sophia S.  
APPLICANT: Napier, Mary A.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Shelton, David B.  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: P2630P1C9  
CURRENT FILING DATE: 2001-10-15  
PRIOR APPLICATION NUMBER: 09/918585  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/064249  
PRIOR FILING DATE: 1997-11-03  
PRIOR APPLICATION NUMBER: 60/065311  
PRIOR FILING DATE: 1997-11-13  
PRIOR APPLICATION NUMBER: 60/066364  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 60/077450  
PRIOR FILING DATE: 1998-03-10  
PRIOR APPLICATION NUMBER: 60/077632  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077641  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077649  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077791  
PRIOR FILING DATE: 1998-03-12  
PRIOR APPLICATION NUMBER: 60/078004  
PRIOR FILING DATE: 1998-03-13  
PRIOR APPLICATION NUMBER: 60/078886  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078936  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078910  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078939  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079294  
PRIOR FILING DATE: 1998-03-25  
PRIOR APPLICATION NUMBER: 60/079656  
PRIOR FILING DATE: 1998-03-26  
PRIOR APPLICATION NUMBER: 60/079664  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079689  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079663  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079728  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079786  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079920  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/079923  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/080105  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080107  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080165  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080194  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080327  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080328  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080333  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080334  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/081070  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081049  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081071  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081195  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081203  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081229  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081955  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081817  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081819  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081952  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081838  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/082568  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082569  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082704  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082804  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082700  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082797  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082796  
PRIOR FILING DATE: 1998-04-23  
PRIOR APPLICATION NUMBER: 60/083336  
PRIOR FILING DATE: 1998-04-27  
PRIOR APPLICATION NUMBER: 60/083322  
PRIOR FILING DATE: 1998-04-28  
PRIOR APPLICATION NUMBER: 60/083392  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083495  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083496  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083499  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083545  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083554

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 18

```

/ PRIOR FILING DATE: 1998-04-29
/ PRIOR APPLICATION NUMBER: 60/083558
/ PRIOR FILING DATE: 1998-04-29
/ PRIOR APPLICATION NUMBER: 60/083559
/ PRIOR FILING DATE: 1998-04-29
/ PRIOR APPLICATION NUMBER: 60/083500
/ PRIOR FILING DATE: 1998-04-29
/ PRIOR APPLICATION NUMBER: 60/083742
/ PRIOR FILING DATE: 1998-04-30
/ PRIOR APPLICATION NUMBER: 60/084366
/ PRIOR FILING DATE: 1998-05-05
/ PRIOR APPLICATION NUMBER: 60/084414
/ PRIOR FILING DATE: 1998-05-06
/ PRIOR APPLICATION NUMBER: 60/084441
/ PRIOR FILING DATE: 1998-05-06
/ PRIOR APPLICATION NUMBER: 60/084637
/ PRIOR FILING DATE: 1998-05-07
/ PRIOR APPLICATION NUMBER: 60/084639
/ PRIOR FILING DATE: 1998-05-07
/ PRIOR APPLICATION NUMBER: 60/084640
/ PRIOR FILING DATE: 1998-05-07
/ PRIOR APPLICATION NUMBER: 60/084598
/ PRIOR FILING DATE: 1998-05-07
/ PRIOR APPLICATION NUMBER: 60/084600
/ PRIOR FILING DATE: 1998-05-07
/ PRIOR APPLICATION NUMBER: 60/084627
/ PRIOR FILING DATE: 1998-05-07
/ PRIOR APPLICATION NUMBER: 60/084643
/ PRIOR FILING DATE: 1998-05-07
/ PRIOR APPLICATION NUMBER: 60/085339
/ PRIOR FILING DATE: 1998-05-13
/ PRIOR APPLICATION NUMBER: 60/085338
/ PRIOR FILING DATE: 1998-05-13
/ PRIOR APPLICATION NUMBER: 60/085323
/ PRIOR FILING DATE: 1998-05-13
/ PRIOR APPLICATION NUMBER: 60/085582
/ PRIOR FILING DATE: 1998-05-15
/ PRIOR APPLICATION NUMBER: 60/085700
/ PRIOR FILING DATE: 1998-05-15
/ PRIOR APPLICATION NUMBER: 60/085689
/ PRIOR FILING DATE: 1998-05-15
/ PRIOR APPLICATION NUMBER: 60/085579
/ PRIOR FILING DATE: 1998-05-15
/ PRIOR APPLICATION NUMBER: 60/085580
/ PRIOR FILING DATE: 1998-05-15
/ PRIOR APPLICATION NUMBER: 60/085573
/ PRIOR FILING DATE: 1998-05-15
/ PRIOR APPLICATION NUMBER: 60/085704
/ PRIOR FILING DATE: 1998-05-15
/ PRIOR APPLICATION NUMBER: 60/085697

```

Query Match 4.3%; Score 80.5; DB 9; Length 736;  
Best Local Similarity 23.8%; Pred. No. 11;  
Matches 30; Conservative 22; Mismatches 41; Indels 33; Gaps 5;

```

Qy 197 DLHDLGHVPMILHP-----SFSEFFINNGRPF--TVI-EKVOALPSKKQ 240
Db 464 DAYDWIGFPDFLEKELDDVYDGEIYSEDSFQNMNLNYSASAVVMDQARKPPSRDQ 523
Qy 241 RIOTLGSNLAIVRCFWFVSEGL-----ENHEGRKAYGAVLISPOELGHA 288
Db 524 WSWTFPQT-----VNAVYLPYTKNEIVFPAGILQAFYARAHNPXKALNFGGIGVMGHELTTHA 578
Qy 289 FIDNVR 294
Db 579 FDDQGR 584

```

RESULT 32  
US-09-999-832a-526  
; Sequence 526, Application US/09999832A  
; Publication No. US20020192706A1  
; GENERAL INFORMATION:

```

/ APPLICANT: Asphenazi, Avi
/ APPLICANT: Baker Kevin P.
/ APPLICANT: Botstein, David
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Eaton, Dan
/ APPLICANT: Ferrara, Napoleon
/ APPLICANT: Filvaroff, Ellen
/ APPLICANT: Fong, Sherman
/ APPLICANT: Gao, Wei-Qiang
/ APPLICANT: Gerber, Hanspeter
/ APPLICANT: Gertlisen, Mary E.
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Grimaldi, J. Christopher
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Hillan, Kenneth J.
/ APPLICANT: Kljavin, Ivar J.
/ APPLICANT: Kuo, Sophia S.
/ APPLICANT: Napier, Mary A.
/ APPLICANT: Pan, James
/ APPLICANT: Pami, Nicholas P.
/ APPLICANT: Roy, Margaret Ann
/ APPLICANT: Shelton, David L.
/ APPLICANT: Stewart, Timothy A.
/ APPLICANT: Tumas, Daniel
/ APPLICANT: Williams, P. Mickey
/ APPLICANT: Wood, William I.
/ TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
/ FILE REFERENCE: P2630PIC63
/ CURRENT FILING DATE: 2001-10-24
/ PRIOR APPLICATION NUMBER: 09/918585
/ PRIOR FILING DATE: 2001-07-30
/ PRIOR APPLICATION NUMBER: 60/062250
/ PRIOR FILING DATE: 1997-10-17
/ PRIOR APPLICATION NUMBER: 60/064249
/ PRIOR FILING DATE: 1997-11-03
/ PRIOR APPLICATION NUMBER: 60/065311
/ PRIOR FILING DATE: 1997-11-13
/ PRIOR APPLICATION NUMBER: 60/066364
/ PRIOR FILING DATE: 1997-11-21
/ PRIOR APPLICATION NUMBER: 60/077450
/ PRIOR FILING DATE: 1998-03-10
/ PRIOR APPLICATION NUMBER: 60/077632
/ PRIOR FILING DATE: 1998-03-11
/ PRIOR APPLICATION NUMBER: 60/077641
/ PRIOR FILING DATE: 1998-03-11
/ PRIOR APPLICATION NUMBER: 60/077649
/ PRIOR FILING DATE: 1998-03-11
/ PRIOR APPLICATION NUMBER: 60/077791
/ PRIOR FILING DATE: 1998-03-12
/ PRIOR APPLICATION NUMBER: 60/078004
/ PRIOR FILING DATE: 1998-03-13
/ PRIOR APPLICATION NUMBER: 60/078886
/ PRIOR FILING DATE: 1998-03-20
/ PRIOR APPLICATION NUMBER: 60/078936
/ PRIOR FILING DATE: 1998-03-20
/ PRIOR APPLICATION NUMBER: 60/078910
/ PRIOR FILING DATE: 1998-03-20
/ PRIOR APPLICATION NUMBER: 60/078939
/ PRIOR FILING DATE: 1998-03-20
/ PRIOR APPLICATION NUMBER: 60/079294
/ PRIOR FILING DATE: 1998-03-25
/ PRIOR APPLICATION NUMBER: 60/079656
/ PRIOR FILING DATE: 1998-03-26
/ PRIOR APPLICATION NUMBER: 60/079664
/ PRIOR FILING DATE: 1998-03-27
/ PRIOR APPLICATION NUMBER: 60/079689
/ PRIOR FILING DATE: 1998-03-27
/ PRIOR APPLICATION NUMBER: 60/079663
/ PRIOR FILING DATE: 1998-03-27
/ PRIOR APPLICATION NUMBER: 60/079728

```

us-09-438-185a-1047.rapb

Fri Jan 10 10:56:11 2003

;; PRIOR FILING DATE: 1998-03-27  
;; PRIOR APPLICATION NUMBER: 60/079786  
;; PRIOR FILING DATE: 1998-03-27  
;; PRIOR APPLICATION NUMBER: 60/079920  
;; PRIOR FILING DATE: 1998-03-30  
;; PRIOR APPLICATION NUMBER: 60/079923  
;; PRIOR FILING DATE: 1998-03-30  
;; PRIOR APPLICATION NUMBER: 60/080105  
;; PRIOR FILING DATE: 1998-03-31  
;; PRIOR APPLICATION NUMBER: 60/080107  
;; PRIOR FILING DATE: 1998-03-31  
;; PRIOR APPLICATION NUMBER: 60/080165  
;; PRIOR FILING DATE: 1998-03-31  
;; PRIOR APPLICATION NUMBER: 60/080194  
;; PRIOR FILING DATE: 1998-03-31  
;; PRIOR APPLICATION NUMBER: 60/080327  
;; PRIOR FILING DATE: 1998-04-01  
;; PRIOR APPLICATION NUMBER: 60/080328  
;; PRIOR FILING DATE: 1998-04-01  
;; PRIOR APPLICATION NUMBER: 60/080333  
;; PRIOR FILING DATE: 1998-04-01  
;; PRIOR APPLICATION NUMBER: 60/080334  
;; PRIOR FILING DATE: 1998-04-01  
;; PRIOR APPLICATION NUMBER: 60/081070  
;; PRIOR FILING DATE: 1998-04-08  
;; PRIOR APPLICATION NUMBER: 60/081049  
;; PRIOR FILING DATE: 1998-04-08  
;; PRIOR APPLICATION NUMBER: 60/081071  
;; PRIOR FILING DATE: 1998-04-08  
;; PRIOR APPLICATION NUMBER: 60/081195  
;; PRIOR FILING DATE: 1998-04-08  
;; PRIOR APPLICATION NUMBER: 60/081203  
;; PRIOR FILING DATE: 1998-04-09  
;; PRIOR APPLICATION NUMBER: 60/081229  
;; PRIOR FILING DATE: 1998-04-09  
;; PRIOR APPLICATION NUMBER: 60/081955  
;; PRIOR FILING DATE: 1998-04-15  
;; PRIOR APPLICATION NUMBER: 60/081817  
;; PRIOR FILING DATE: 1998-04-15  
;; PRIOR APPLICATION NUMBER: 60/081819  
;; PRIOR FILING DATE: 1998-04-15  
;; PRIOR APPLICATION NUMBER: 60/081952  
;; PRIOR FILING DATE: 1998-04-15  
;; PRIOR APPLICATION NUMBER: 60/081838  
;; PRIOR FILING DATE: 1998-04-15  
;; PRIOR APPLICATION NUMBER: 60/082568  
;; PRIOR FILING DATE: 1998-04-21  
;; PRIOR APPLICATION NUMBER: 60/082569  
;; PRIOR FILING DATE: 1998-04-21  
;; PRIOR APPLICATION NUMBER: 60/082704  
;; PRIOR FILING DATE: 1998-04-22  
;; PRIOR APPLICATION NUMBER: 60/082804  
;; PRIOR FILING DATE: 1998-04-22  
;; PRIOR APPLICATION NUMBER: 60/082700  
;; PRIOR FILING DATE: 1998-04-22  
;; PRIOR APPLICATION NUMBER: 60/082797  
;; PRIOR FILING DATE: 1998-04-22  
;; PRIOR APPLICATION NUMBER: 60/082796  
;; PRIOR FILING DATE: 1998-04-23  
;; PRIOR APPLICATION NUMBER: 60/083336  
;; PRIOR FILING DATE: 1998-04-27  
;; PRIOR APPLICATION NUMBER: 60/083322  
;; PRIOR FILING DATE: 1998-04-28  
;; PRIOR APPLICATION NUMBER: 60/083392  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083495  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083496  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083499  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083545  
;; PRIOR FILING DATE: 1998-04-29

;; PRIOR APPLICATION NUMBER: 60/083554  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083558  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083559  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083500  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083742  
;; PRIOR FILING DATE: 1998-04-30  
;; PRIOR APPLICATION NUMBER: 60/084366  
;; PRIOR FILING DATE: 1998-05-05  
;; PRIOR APPLICATION NUMBER: 60/084414  
;; PRIOR FILING DATE: 1998-05-06  
;; PRIOR APPLICATION NUMBER: 60/084441  
;; PRIOR FILING DATE: 1998-05-06  
;; PRIOR APPLICATION NUMBER: 60/084637  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084639  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084640  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084598  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084600  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084627  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084643  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/085339  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: 60/085338  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: 60/085323  
;; PRIOR FILING DATE: 1998-05-13  
;; PRIOR APPLICATION NUMBER: 60/085582  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085700  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085689  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085579  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085580  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085573  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085704  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085697

Query Match 4.3%; Score 80.5; DB 9; Length 736;  
Best Local Similarity 23.8%; Pred. No. 11;  
Matches 30; Conservative 22; Mismatches 41; Indels 33; Gaps 5;  
Qy 197 DIHLLGHVPWLLHP-----SFSEFFINMGRLP---TKVI-EKVQALPSKKQ 240  
Db 464 DAYDMIGPDPFILEPEKELDDVDGYEISEDFFQNNLNLNFSAKVMADQLRKPPSRDQ 523  
Qy 241 RIQTQSNLIAIVRCFMFTVESGLI-----ENHEGRKAYGAVLISSPQELGHA 288  
Db 524 WSMTPOT-----VNAYLPTKNEIVFPAGILQAPFYARNHPKALNFGGIGVVMGHELTHA 578  
Qy 289 FIDNVR 294  
Db 579 FDDQGR 584

RESULT 33  
US-09-978-189-526  
; Sequence 526, Application US/09978189  
; Publication No. US20030004102A1

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 20

GENERAL INFORMATION:

APPLICANT: Ashkenazi, Avi  
APPLICANT: Baker Kevin P.  
APPLICANT: Botstein, David  
APPLICANT: Desnovers, Luc  
APPLICANT: Eaton, Dan  
APPLICANT: Ferrara, Napoleon  
APPLICANT: Filvarolt, Ellen  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Geritsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, J. Christopher  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth J.  
APPLICANT: Kijavlin, Ivar J.  
APPLICANT: Kuo, Sophia S.  
APPLICANT: Napier, Mary A.  
APPLICANT: Pan, James.  
APPLICANT: Beon, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Shelton, David L.  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: P2630PLC7  
CURRENT APPLICATION NUMBER: US/09/978,189  
PRIOR FILING DATE: 2001-10-15  
PRIOR APPLICATION NUMBER: 09/918585  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/064249  
PRIOR FILING DATE: 1997-11-03  
PRIOR APPLICATION NUMBER: 60/065311  
PRIOR FILING DATE: 1997-11-13  
PRIOR APPLICATION NUMBER: 60/066364  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 60/074450  
PRIOR FILING DATE: 1998-03-10  
PRIOR APPLICATION NUMBER: 60/077632  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077641  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077649  
PRIOR FILING DATE: 1998-03-11  
PRIOR APPLICATION NUMBER: 60/077791  
PRIOR FILING DATE: 1998-03-12  
PRIOR APPLICATION NUMBER: 60/078004  
PRIOR FILING DATE: 1998-03-13  
PRIOR APPLICATION NUMBER: 60/078886  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078936  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078910  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/078939  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079294  
PRIOR FILING DATE: 1998-03-25  
PRIOR APPLICATION NUMBER: 60/079656  
PRIOR FILING DATE: 1998-03-26  
PRIOR APPLICATION NUMBER: 60/079664  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079689  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079663  
PRIOR FILING DATE: 1998-03-27

PRIOR APPLICATION NUMBER: 60/079728  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079786  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079920  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/079923  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/080105  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080107  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080165  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080194  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080327  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080328  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080333  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080334  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/081070  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081049  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081071  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081195  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081203  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081229  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081955  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081817  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081819  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081952  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081838  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/082568  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082569  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082704  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082804  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082700  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082797  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082796  
PRIOR FILING DATE: 1998-04-23  
PRIOR APPLICATION NUMBER: 60/083336  
PRIOR FILING DATE: 1998-04-27  
PRIOR APPLICATION NUMBER: 60/083322  
PRIOR FILING DATE: 1998-04-28  
PRIOR APPLICATION NUMBER: 60/083392  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083495  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083496  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083499  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083545

us-09-438-185a-1047.rapb

Fri Jan 10 10:56:11 2003

```

PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083554
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083558
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083500
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083742
PRIOR FILING DATE: 1998-04-30
PRIOR APPLICATION NUMBER: 60/084366
PRIOR FILING DATE: 1998-05-05
PRIOR APPLICATION NUMBER: 60/084414
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084441
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084639
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084640
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084598
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084627
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084643
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085339
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085338
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085582
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085700
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085689
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085580
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085573
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085704
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085697

Query Match 4.3%; Score 80.5; DB 9; Length 736;
Best Local Similarity 23.8%; Pred. No. 11;
Matches 30; Conservative 22; Mismatches 41; Indels 33; Gaps 5;

QY 197 DLHDLGHVFWLLHP-----SFSEFFINMGRLF---TKVI-EKVOALPSKKQ 240
DB 464 DAIVDMIGFPDFILEKPELDDVDYDGYEISDSFFQNNMLNLYNFSAKVMADQURKPPSRDQ 523
QY 241 RIQTLQNLIAIVRCFWFVVEGLI-----ENHEGRKAYGAVLISSPQELGHA 288
DB 524 WSMTPQT-----VNAYVLPKNEIVFPAGILQAPFYARNHPKALNFGGIVGVVMGHELTHA 578
QY 289 FIDNVR 294
DB 579 FDDQGR 584

RESULT 34
US-10-052-586-420
; Sequence 420, Application US/10052586

```

```

Patent No. US20020127584A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zhen
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C1
CURRENT APPLICATION NUMBER: US/10/052,586
CURRENT FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063564
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063734
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063870
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/064103
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066120
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/066466
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066772
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/069335
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069425
PRIOR FILING DATE: 1997-12-12
PRIOR APPLICATION NUMBER: 60/069870
PRIOR FILING DATE: 1997-12-17
PRIOR APPLICATION NUMBER: 60/068017
PRIOR FILING DATE: 1997-12-18
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786

```

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 22

PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/080107  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080194  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/080327  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/080333  
PRIOR FILING DATE: 1998-04-01  
PRIOR APPLICATION NUMBER: 60/081049  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081070  
PRIOR FILING DATE: 1998-04-08  
PRIOR APPLICATION NUMBER: 60/081195  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081838  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/082568  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082569  
PRIOR FILING DATE: 1998-04-21  
PRIOR APPLICATION NUMBER: 60/082704  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/082797  
PRIOR FILING DATE: 1998-04-22  
PRIOR APPLICATION NUMBER: 60/083322  
PRIOR FILING DATE: 1998-04-28  
PRIOR APPLICATION NUMBER: 60/083495  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083496  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083499  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/083559  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/084366  
PRIOR FILING DATE: 1998-05-05  
PRIOR APPLICATION NUMBER: 60/084414  
PRIOR FILING DATE: 1998-05-06  
PRIOR APPLICATION NUMBER: 60/084639  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084640  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084643  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/085573  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085579  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085580  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085582  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085700  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/086023  
PRIOR FILING DATE: 1998-05-18  
PRIOR APPLICATION NUMBER: 60/086392  
PRIOR FILING DATE: 1998-05-22  
PRIOR APPLICATION NUMBER: 60/086486  
PRIOR FILING DATE: 1998-05-22  
PRIOR APPLICATION NUMBER: 60/087098  
PRIOR FILING DATE: 1998-05-28  
PRIOR APPLICATION NUMBER: 60/087208  
PRIOR FILING DATE: 1998-05-28  
PRIOR APPLICATION NUMBER: 60/087609  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/087759  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/087827  
PRIOR FILING DATE: 1998-06-03  
PRIOR APPLICATION NUMBER: 60/088025  
PRIOR FILING DATE: 1998-06-04

PRIOR APPLICATION NUMBER: 60/088028  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088029  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088033  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088167  
PRIOR FILING DATE: 1998-06-05  
PRIOR APPLICATION NUMBER: 60/088202  
PRIOR FILING DATE: 1998-06-05  
PRIOR APPLICATION NUMBER: 60/088212  
PRIOR FILING DATE: 1998-06-05  
PRIOR APPLICATION NUMBER: 60/088217  
PRIOR FILING DATE: 1998-06-05  
PRIOR APPLICATION NUMBER: 60/088326  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088655  
PRIOR FILING DATE: 1998-06-09  
PRIOR APPLICATION NUMBER: 60/088722  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088738  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088740  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088811  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088824  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088825  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088826  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088861  
PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/088863  
PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/088876  
PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/089090  
PRIOR FILING DATE: 1998-06-12  
PRIOR APPLICATION NUMBER: 60/089105  
PRIOR FILING DATE: 1998-06-12  
PRIOR APPLICATION NUMBER: 60/089512  
PRIOR FILING DATE: 1998-06-16  
PRIOR APPLICATION NUMBER: 60/089514  
PRIOR FILING DATE: 1998-06-16  
PRIOR APPLICATION NUMBER: 60/089538  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089598  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089653  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089908

Query Match 4.3%; Score 80.5; DB 12; Length 736;  
Best Local Similarity 23.8%; Pred. No. 11;  
Matches 30; Conservative 22; Mismatches 41; Indels 33; Caps 5;

QY 137 DLHDLGHVPLHLP-----SFSEFIMNGRPF--TKVI-EKVOALPSKQ 240  
DB 444 DAIIDMIGFPDFILFKELDDYDGYEISBDSFFONMLNINFSAKYMAOQKPPSRQ 523  
QY 241 RIQLOGLNLAIVRCFWTFVESGLI-----ENHGRKAYGAVLISSPQELGHA 288  
DB 524 WSMPTQT-----VNAYYLPKNEIVFPAGLLOAPFYARNHPKALNFGIGVVGHELTTHA 578  
QY 289 FIDNVR 294  
DB 579 FDDQGR 584

RESULT 35  
US-09-888-615-80



```
; Sequence 80, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: WHYTE, DAVID
; APPLICANT: CAENEPEEL, SEAN
; APPLICANT: CHARYDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: US/09/898,615
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,047
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 80
; LENGTH: 765
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-888-615-80

Query Match      4.3%; Score 80.5; DB 10; Length 765;
Best Local Similarity 23.8%; Pred. No. 11; Indels 33; Gaps 5;
Matches 30; Conservative 22; Mismatches 41;

QY 197 DLHDLGHVPLWLP-----SFSEFFINMGRLP---TKVI-EKVQALPSKKQ 240
DB 493 DAIYDMIGFPDFFILEKELDDVDYGVIESEDSFFQNMNLNLYNFAKAWAQLRKPPSRQ 552

QY 241 RIQTQSNLIAIVRCFWFTVESGLI-----ENHGRKAYGAVLISSPQELGHA 288
DB 553 WSMTPQT-----VNAYLPTKNEIVFPAGILQAPFYARNHPKALNFGGIGVVMGHETHA 607

QY 289 FIDNVR 294
DB 608 FDDQGR 613

RESULT 36
US-09-815-242-13304
; Sequence 13304, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11609
; LENGTH: 897
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; US-09-815-242-11609

Query Match      4.2%; Score 80; DB 10; Length 897;
```

Best Local Similarity 19.5%; Pred. No. 16;  
Matches 59; Conservative 54; Mismatches 104; Indels 86; Gaps 14;  
QY 27 PFONSQSLQAYSTPYSYRIILOKENEKQALARK----- 63  
Db 54 FYKDKRM-----PFIVALSQRTKRAELGEYKONRKADAKEMLOIPALEMLQK 107  
QY 64 ---CISLEFFKMLLFVHLISLKNQREGSTMAVVSPPFRNIMWRLSSRSLMK 119  
Db 108 MGTCEVGGFEADVDIASLATLSPKRIYSKXD-----FK-----QLSDKIALPD 156  
QY 120 SYCPREFLDYLAAGLL-SDPLDQAVIKFPELETHPSYVSGFVAHQYLSLQDRYF 178  
Db 157 GKTFFLAKDCVEKYGILPSQFTDYOGIVG---DSDNYKGVKG-IGSKNAKELLQ-RLG 210  
QY 179 PIASVMTLDKDNFSITPDILHDLGHVWMLHPFSSEFFINMGRLEFTKVIKYOALPSK 238  
Db 211 SLEKIVENLDLAKNLSPPWYOAL-----IQDKGSAFLSK 245  
QY 239 KQRIQTOSSLIA---IVRCFWPTVESGLIENHGRKAYGAVILSSPOLGHA--PIDNV 293  
Db 246 E--LAIHERGCIKEPFLSCA-FPSENPLIKIDELKYG--FISTLRDLNRPFIYENV 300  
QY 294 RVL 296  
Db 301 PTL 303  
RESULT 38  
US-09-731-231A-2  
Sequence 2, Application US/09731231A  
Patent No. US20020082189A1  
GENERAL INFORMATION:  
APPLICANT: GUEGLER, Karl et al  
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
TITLE OF INVENTION: THERMOF  
FILE REFERENCE: CL001007  
CURRENT APPLICATION NUMBER: US/09/731,231A  
NUMBER OF SEQ. ID NOS: 6  
CURRENT FILING DATE: 2000-12-07  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 645  
TYPE: PRT  
ORGANISM: Human  
US-09-731-231A-2

Query Match 4.2%; Score 79; DB 10; Length 645;  
Best Local Similarity 19.7%; Pred. No. 13;  
Matches 55; Conservative 47; Mismatches 107; Indels 70; Gaps 12;

QY 3 YCE-RTLDPYIILKALROSLSLFQNI---SGLQRAVSTPYRIILOKENEKQ 58  
Db 110 YCEGRDLDDK-----IOEYKQAGKIFPENQIIEWPIQLLGVDYMERRLIHRDLKSNV 164  
QY 59 LARHKCISLEFFKMLLFVHLISLKNQREGSTMAVVSPPFR--NRNMYRLSSRS 116  
Db 165 FLKKNLAKIGDF-----GVRILMGSCDLATLTGTGPHMSPEALKHGYOTKSD 214  
QY 117 LMKSYCPREFLDYLAAGLLSDPLDQAVIKFPELETHPSYVSGFVAHQYLSLQDR 176  
Db 215 IWSLACILYEMCCMHAFAGSNPL--STVLKIVESDT-----PSLPER 255  
QY 177 Y-FPIASVMTLDKDNFSITPDILHDLGHVWML-----LHPSFSEFFINMGRLE 224  
Db 256 YPKELAINMESMLNKNPSLRPSAIEL--KIYVDEOQLNMCVSEMTLEDKNLDCQRE 313  
QY 225 -----FTKVIKYOAL-PSKQRIQTOSSL 247  
Db 314 AAHINAMOKRIHLQTLRALISEVQKTPRERMRKLQDA 352

RESULT 39  
US-09-992-598-84  
Sequence 84, Application US/09992598  
Patent No. US20020160384A1  
GENERAL INFORMATION:  
APPLICANT: Ashkenazi, Avi J.  
APPLICANT: Baker, Kevin P.  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan L.  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Fong, Sherman  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gerlitsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, J. Christopher  
APPLICANT: Gurney, Austin L.  
APPLICANT: Kijavyn, Ivar J.  
APPLICANT: Napiet, Mary A.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
TITLE OF INVENTION: Acids Encoding the Same  
FILE REFERENCE: P2730PIC20  
CURRENT APPLICATION NUMBER: US/09/992,598  
PRIOR FILING DATE: 2001-11-14  
PRIOR APPLICATION NUMBER: 60/049787  
PRIOR FILING DATE: 1997-06-16  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/065186  
PRIOR FILING DATE: 1997-11-12  
PRIOR APPLICATION NUMBER: 60/065311  
PRIOR FILING DATE: 1997-11-13  
PRIOR APPLICATION NUMBER: 60/066770  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/075945  
PRIOR FILING DATE: 1998-02-25  
PRIOR APPLICATION NUMBER: 60/078910  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/083322  
PRIOR FILING DATE: 1998-04-28  
PRIOR APPLICATION NUMBER: 60/084600  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/087106  
PRIOR FILING DATE: 1998-05-28  
PRIOR APPLICATION NUMBER: 60/087607  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/087609  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/087759  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/087827  
PRIOR FILING DATE: 1998-06-03  
PRIOR APPLICATION NUMBER: 60/088021  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088025  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088026  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088028  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088029  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088030

us-09-438-185a-1047.rapb

Fri Jan 10 10:56:11 2003

;;  
;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088033  
;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088326  
;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088167  
;; PRIOR FILING DATE: 1998-06-05  
;; PRIOR APPLICATION NUMBER: 60/088202  
;; PRIOR FILING DATE: 1998-06-05  
;; PRIOR APPLICATION NUMBER: 60/088212  
;; PRIOR FILING DATE: 1998-06-05  
;; PRIOR APPLICATION NUMBER: 60/088217  
;; PRIOR FILING DATE: 1998-06-05  
;; PRIOR APPLICATION NUMBER: 60/088655  
;; PRIOR FILING DATE: 1998-06-09  
;; PRIOR APPLICATION NUMBER: 60/088734  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088738  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088742  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088810  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088824  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088826  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088858  
;; PRIOR FILING DATE: 1998-06-11  
;; PRIOR APPLICATION NUMBER: 60/088861  
;; PRIOR FILING DATE: 1998-06-11  
;; PRIOR APPLICATION NUMBER: 60/088876  
;; PRIOR FILING DATE: 1998-06-11  
;; PRIOR APPLICATION NUMBER: 60/089105  
;; PRIOR FILING DATE: 1998-06-12  
;; PRIOR APPLICATION NUMBER: 60/089440  
;; PRIOR FILING DATE: 1998-06-16  
;; PRIOR APPLICATION NUMBER: 60/089512  
;; PRIOR FILING DATE: 1998-06-16  
;; PRIOR APPLICATION NUMBER: 60/089514  
;; PRIOR FILING DATE: 1998-06-16  
;; PRIOR APPLICATION NUMBER: 60/089532  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089538  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089598  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089599  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089600  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089653  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089801  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: 60/089907  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: 60/089908  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: 60/089947  
;; PRIOR FILING DATE: 1998-06-19  
;; PRIOR APPLICATION NUMBER: 60/089948  
;; PRIOR FILING DATE: 1998-06-19  
;; PRIOR APPLICATION NUMBER: 60/089952  
;; PRIOR FILING DATE: 1998-06-19  
;; PRIOR APPLICATION NUMBER: 60/090246  
;; PRIOR FILING DATE: 1998-06-22  
;; PRIOR APPLICATION NUMBER: 60/090252  
;; PRIOR FILING DATE: 1998-06-22  
;; PRIOR APPLICATION NUMBER: 60/090254  
;; PRIOR FILING DATE: 1998-06-22  
;; PRIOR APPLICATION NUMBER: 60/090349  
;; PRIOR FILING DATE: 1998-06-23

;; PRIOR APPLICATION NUMBER: 60/090355  
;; PRIOR FILING DATE: 1998-06-23  
;; PRIOR APPLICATION NUMBER: 60/090429  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090431  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090435  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090444  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090445  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090472  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090535  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090540  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090542  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090557  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090676  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090678  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090690  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090694  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090695  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090696  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/090862  
;; PRIOR FILING DATE: 1998-06-26  
;; PRIOR APPLICATION NUMBER: 60/090863  
;; PRIOR FILING DATE: 1998-06-26  
;; PRIOR APPLICATION NUMBER: 60/091360  
;; PRIOR FILING DATE: 1998-07-01  
;; PRIOR APPLICATION NUMBER: 60/091478  
;; PRIOR FILING DATE: 1998-07-02  
;; PRIOR APPLICATION NUMBER: 60/091544  
;; PRIOR FILING DATE: 1998-07-01  
;; PRIOR APPLICATION NUMBER: 60/091519  
;; PRIOR FILING DATE: 1998-07-02  
;; PRIOR APPLICATION NUMBER: 60/091626  
;; PRIOR FILING DATE: 1998-07-02  
;; PRIOR APPLICATION NUMBER: 60/091633  
;; PRIOR FILING DATE: 1998-07-02  
;; PRIOR APPLICATION NUMBER: 60/091978  
;; PRIOR FILING DATE: 1998-07-07  
;; PRIOR APPLICATION NUMBER: 60/091982  
;; PRIOR FILING DATE: 1998-07-07  
;; PRIOR APPLICATION NUMBER: 60/092182  
;; PRIOR FILING DATE: 1998-07-09

Query Match 4.2%; Score 79; DB 9; Length 867;  
Best Local Similarity 23.8%; Pred. No. 19;  
Matches 59; Conservative 34; Mismatches 85; Indels 70; Gaps 16;

Qy 99 TPFENRL-----WYRLI-SGRFSLWKSVC-----PRFFLDYLEAFGLSD 139  
Db 138 TAFPGKVLNEYNGSVYPPGKWEVGLLKNRSFYNY-TLCRNGVKKGHSKDY--LTD 194  
Qy 139 FLHQAVIKFELEHFSYYP-----VSGVAPH-----QYLSLLQDRYFPFIASVWR 185  
Db 195 LITNDSV-SFF--RTSKNYPHPRVLMVISHAAPHGPDSPQY-----SRLFPNAS-QH 245  
Qy 186 TLDKDNFSLTPDLIHLLGHVPLHPSFSEFFINMGRLETKVIERVQALPSKKORIOTL 245  
Db 246 ITPSYNAPND-----KEWIMRYGPKMKPHME--FTNMLQ-----RKRLQTL 287

; PRIOR APPLICATION NUMBER: 60/088025;  
 ; PRIOR FILING DATE: 1998-06-04  
 ; PRIOR APPLICATION NUMBER: 60/088026;  
 ; PRIOR FILING DATE: 1998-06-04  
 ; PRIOR APPLICATION NUMBER: 60/088028;  
 ; PRIOR FILING DATE: 1998-06-04  
 ; PRIOR APPLICATION NUMBER: 60/088025;

PRIOR	FILING DATE:	1996-06-04
PRIOR	APPLICATION NUMBER:	60/088033
PRIOR	FILING DATE:	1996-06-04
PRIOR	APPLICATION NUMBER:	60/088322
PRIOR	FILING DATE:	1996-06-04
PRIOR	APPLICATION NUMBER:	60/088167
PRIOR	FILING DATE:	1996-06-05
PRIOR	APPLICATION NUMBER:	60/088207
PRIOR	FILING DATE:	1996-06-05
PRIOR	APPLICATION NUMBER:	60/088212
PRIOR	FILING DATE:	1996-06-05
PRIOR	APPLICATION NUMBER:	60/08821
PRIOR	FILING DATE:	1996-06-05
PRIOR	APPLICATION NUMBER:	60/088655
PRIOR	FILING DATE:	1996-06-05
PRIOR	APPLICATION NUMBER:	60/08873
PRIOR	FILING DATE:	1996-06-10
PRIOR	APPLICATION NUMBER:	60/088733
PRIOR	FILING DATE:	1996-06-10
PRIOR	APPLICATION NUMBER:	60/088744
PRIOR	FILING DATE:	1996-06-10
PRIOR	APPLICATION NUMBER:	60/088810
PRIOR	FILING DATE:	1996-06-10
PRIOR	APPLICATION NUMBER:	60/088824
PRIOR	FILING DATE:	1996-06-10
PRIOR	APPLICATION NUMBER:	60/088822
PRIOR	FILING DATE:	1996-06-10
PRIOR	APPLICATION NUMBER:	60/088855
PRIOR	FILING DATE:	1996-06-11
PRIOR	APPLICATION NUMBER:	60/088866
PRIOR	FILING DATE:	1996-06-11
PRIOR	APPLICATION NUMBER:	60/088877
PRIOR	FILING DATE:	1996-06-11
PRIOR	APPLICATION NUMBER:	60/089105
PRIOR	FILING DATE:	1996-06-12
PRIOR	APPLICATION NUMBER:	60/089444
PRIOR	FILING DATE:	1996-06-16
PRIOR	APPLICATION NUMBER:	60/089512
PRIOR	FILING DATE:	1996-06-16
PRIOR	APPLICATION NUMBER:	60/089514
PRIOR	FILING DATE:	1996-06-16
PRIOR	APPLICATION NUMBER:	60/089533
PRIOR	FILING DATE:	1996-06-17
PRIOR	APPLICATION NUMBER:	60/089539
PRIOR	FILING DATE:	1996-06-17
PRIOR	APPLICATION NUMBER:	60/089599
PRIOR	FILING DATE:	1996-06-17
PRIOR	APPLICATION NUMBER:	60/089599
PRIOR	FILING DATE:	1996-06-17
PRIOR	APPLICATION NUMBER:	60/089607
PRIOR	FILING DATE:	1996-06-17
PRIOR	APPLICATION NUMBER:	60/089655
PRIOR	FILING DATE:	1996-06-17
PRIOR	APPLICATION NUMBER:	60/089801
PRIOR	FILING DATE:	1996-06-18
PRIOR	APPLICATION NUMBER:	60/089907
PRIOR	FILING DATE:	1996-06-18
PRIOR	APPLICATION NUMBER:	60/089908
PRIOR	FILING DATE:	1996-06-18
PRIOR	APPLICATION NUMBER:	60/089944
PRIOR	FILING DATE:	1996-06-19
PRIOR	APPLICATION NUMBER:	60/089955
PRIOR	FILING DATE:	1996-06-19

[illegible]

us-09-438-185a-1047.rapb

Fri Jan 10 10:56:11 2003

PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: 60/090246  
PRIOR FILING DATE: 1998-06-22  
PRIOR APPLICATION NUMBER: 60/090252  
PRIOR FILING DATE: 1998-06-22  
PRIOR APPLICATION NUMBER: 60/090254  
PRIOR FILING DATE: 1998-06-22  
PRIOR APPLICATION NUMBER: 60/090349  
PRIOR FILING DATE: 1998-06-23  
PRIOR APPLICATION NUMBER: 60/090355  
PRIOR FILING DATE: 1998-06-23  
PRIOR APPLICATION NUMBER: 60/090429  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090431  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090435  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090444  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090445  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090472  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090535  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090540  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090542  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090557  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090676  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090678  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090690  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090694  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090695  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090696  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090862  
PRIOR FILING DATE: 1998-06-26  
PRIOR APPLICATION NUMBER: 60/090863  
PRIOR FILING DATE: 1998-06-26  
PRIOR APPLICATION NUMBER: 60/091360  
PRIOR FILING DATE: 1998-07-01  
PRIOR APPLICATION NUMBER: 60/091478  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091544  
PRIOR FILING DATE: 1998-07-01  
PRIOR APPLICATION NUMBER: 60/091519  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091626  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091633  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091978  
PRIOR FILING DATE: 1998-07-07  
PRIOR APPLICATION NUMBER: 60/091982  
PRIOR FILING DATE: 1998-07-07  
PRIOR APPLICATION NUMBER: 60/092182  
PRIOR FILING DATE: 1998-07-09

Query Match 4.2%; Score 79; DB 9; Length 867;  
Best Local Similarity 23.8%; Pred. No. 19;

Matches 59; Conservative 34; Mismatches 85; Indels 70; Gaps 16;

QY 99 TPFNRM-----WYLL-SRPSLWKSVC-----PRFFLDYLEAFLGSLD 138

DB 138 TAFPGKLYNEINGSYPPGCKEWGLLNRSFYNY-TLCRNGVKXKHSYKDY--LTD 194

QY 139 FLDHQAVIKFPELETHFSYYP-----VSGFVAPH-----QYLSLLQDRYFFPIASVMR 185  
DB 195 LIITDSV-SFF--RTSKMYPHRVLMLVISHAAPHGPDSPAQY-----SRLEFFNAS-QH 245  
QY 186 TLBDKNFSLTDPDLIHDLLGHVPLLLHPSSEFFINNRLFTKVIKVOALPSKKORIOTL 245  
DB 246 ITSINYAANPD-----KHMIRYTGPMKPIHME--FTNMLQ-----RKRLQTL 287  
QY 246 QSNLIAIVRCFWFTVESGLIENHEGRKAYGAVLISSPQELGHAFIDNVRLVPLPLELDQIIR 305  
DB 288 MSVDDSMETIYNMLVETGELDN-----TYIVYTADHGYHIGQGLVKGKSMFYEPD--IR 340  
QY 306 LPFNTSTP 313  
DB 341 VPFYVRGP 348  
RESULT 41  
US-09-989-735-84  
; Sequence 84, Application US/09989735  
; Publication No. US20020193299A1  
; GENERAL INFORMATION:  
; APPLICANT: Ashkenazi, Avi J.  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Botstein, David  
; APPLICANT: Desnovers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, J. Christopher  
; APPLICANT: Guiney, Austin J.  
; APPLICANT: Kijavirin, Ivay J.  
; APPLICANT: Napier, Mary A.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tamas, Daniel  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; TITLE OF INVENTION: Acids Encoding the Same  
; FILE REFERENCE: P2730P1C61  
; CURRENT APPLICATION NUMBER: US/09/989,735  
; CURRENT FILING DATE: 2001-11-19  
; PRIOR APPLICATION NUMBER: 60/049787  
; PRIOR FILING DATE: 1997-06-16  
; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/065186  
; PRIOR FILING DATE: 1997-11-12  
; PRIOR APPLICATION NUMBER: 60/065311  
; PRIOR FILING DATE: 1997-11-13  
; PRIOR APPLICATION NUMBER: 60/066770  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/075945  
; PRIOR FILING DATE: 1998-02-25  
; PRIOR APPLICATION NUMBER: 60/078910  
; PRIOR FILING DATE: 1998-03-20  
; PRIOR APPLICATION NUMBER: 60/083322  
; PRIOR FILING DATE: 1998-04-28  
; PRIOR APPLICATION NUMBER: 60/084600  
; PRIOR FILING DATE: 1998-05-07  
; PRIOR APPLICATION NUMBER: 60/087106  
; PRIOR FILING DATE: 1998-05-28  
; PRIOR APPLICATION NUMBER: 60/087607

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rabb

Page 28

PRIOR FILING DATE	1998-06-02
PRIOR APPLICATION NUMBER:	60/087609
PRIOR FILING DATE:	1998-06-02
PRIOR APPLICATION NUMBER:	60/087759
PRIOR FILING DATE:	1998-06-02
PRIOR APPLICATION NUMBER:	60/087827
PRIOR FILING DATE:	1998-06-03
PRIOR APPLICATION NUMBER:	60/088021
PRIOR FILING DATE:	1998-06-04
PRIOR APPLICATION NUMBER:	60/088025
PRIOR FILING DATE:	1998-06-04
PRIOR APPLICATION NUMBER:	60/088026
PRIOR FILING DATE:	1998-06-04
PRIOR APPLICATION NUMBER:	60/088028
PRIOR FILING DATE:	1998-06-04
PRIOR APPLICATION NUMBER:	60/088029
PRIOR FILING DATE:	1998-06-04
PRIOR APPLICATION NUMBER:	60/088030
PRIOR FILING DATE:	1998-06-04
PRIOR APPLICATION NUMBER:	60/088033
PRIOR FILING DATE:	1998-06-04
PRIOR APPLICATION NUMBER:	60/088326
PRIOR FILING DATE:	1998-06-04
PRIOR APPLICATION NUMBER:	60/088167
PRIOR FILING DATE:	1998-06-05
PRIOR APPLICATION NUMBER:	60/088202
PRIOR FILING DATE:	1998-06-05
PRIOR APPLICATION NUMBER:	60/088212
PRIOR FILING DATE:	1998-06-05
PRIOR APPLICATION NUMBER:	60/088217
PRIOR FILING DATE:	1998-06-05
PRIOR APPLICATION NUMBER:	60/088655
PRIOR FILING DATE:	1998-06-09
PRIOR APPLICATION NUMBER:	60/088734
PRIOR FILING DATE:	1998-06-10
PRIOR APPLICATION NUMBER:	60/088738
PRIOR FILING DATE:	1998-06-10
PRIOR APPLICATION NUMBER:	60/088742
PRIOR FILING DATE:	1998-06-10
PRIOR APPLICATION NUMBER:	60/088810
PRIOR FILING DATE:	1998-06-10
PRIOR APPLICATION NUMBER:	60/088824
PRIOR FILING DATE:	1998-06-10
PRIOR APPLICATION NUMBER:	60/088826
PRIOR FILING DATE:	1998-06-10
PRIOR APPLICATION NUMBER:	60/088858
PRIOR FILING DATE:	1998-06-11
PRIOR APPLICATION NUMBER:	60/088861
PRIOR FILING DATE:	1998-06-11
PRIOR APPLICATION NUMBER:	60/088876
PRIOR FILING DATE:	1998-06-11
PRIOR APPLICATION NUMBER:	60/089105
PRIOR FILING DATE:	1998-06-12
PRIOR APPLICATION NUMBER:	60/089440
PRIOR FILING DATE:	1998-06-16
PRIOR APPLICATION NUMBER:	60/089512
PRIOR FILING DATE:	1998-06-16
PRIOR APPLICATION NUMBER:	60/089514
PRIOR FILING DATE:	1998-06-16
PRIOR APPLICATION NUMBER:	60/089532
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089538
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089598
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089599
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089600
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089653
PRIOR FILING DATE:	1998-06-17
PRIOR APPLICATION NUMBER:	60/089801
PRIOR FILING DATE:	1998-06-18

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

PRIOR FILING DATE: 1998-07-09

Query Match 4.2%; Score 79; DB 9; Length 867;

Best Local Similarity 23.8%; Pred. No. 19;

Matches 59; Conservative 34; Mismatches 85; Indels 70; Gaps 16;

QY 99 TPFFNNRL-----WYRL--SSRFLWKSYC-----PRFFDYLEAPGLSD 138

Db 138 TAFGKVLNENYGVPPGKEMVGLLKNRFFNY-TLCRNGVKRKHGSDYKDY--LTD 194

QY 139 FLDOAVIKPFLETHFSYYP-----VSGFVAPH-----QVLSLODRYFFPIASVMR 185

Db 195 LITNDSV-SFF--RTSKMYPHRPVLWVISHAAPHGPDSPAQY-----SRLFPNAS-QH 245

QY 186 TLDKNSFLPDLHLHGVPLWLLHPSSEFFNMGRFTKVKIEKVQALPSKKORIOTL 245

Db 246 TTPSNVAPND-----KHWMYTGMPKPIHWE--FTNMLQ-----RKELQTL 287

QY 246 QENLIAVRCFWFTVESGLIENHEGRKAYCAVLISSPQELGHAFIDNVRLPLELDQIIR 305

Db 288 MSVDOSMETIINMLVETGELDN-----TIVYADHGVIHQFGLVKGKMPYFED--IR 340

QY 306 LPFNTSTP 313

Db 341 VPFYVRGP 348

RESULT 42

US-09-990-444-84

Sequence 84, Application US/09990444

Publication No. US20020193300A1

GENERAL INFORMATION:

APPLICANT: Ashkenazi, Avi J.

APPLICANT: Baker, Kevin P.

APPLICANT: Botstein, David

APPLICANT: Desnovers, Luc

APPLICANT: Eaton, Dan L.

APPLICANT: Ferrara, Napoleone

APPLICANT: Fong, Sherman

APPLICANT: Gerber, Hanspeter

APPLICANT: Gerritsen, Mary E.

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Grimaldi, J. Christopher

APPLICANT: Gurney, Austin L.

APPLICANT: Kijavitt, Ivar J.

APPLICANT: Napier, Mary A.

APPLICANT: Pan, James

APPLICANT: Paoni, Nicholas F.

APPLICANT: Roy, Margaret Ann

APPLICANT: Stewart, Timothy A.

APPLICANT: Tumas, Daniel

APPLICANT: Watanabe, Colin K.

APPLICANT: Williams, P. Mickey

APPLICANT: Wood, William I.

APPLICANT: Zhang, Zemin

TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

FILE REFERENCE: P2730PIC19

CURRENT APPLICATION NUMBER: US/09/990,444

CURRENT FILING DATE: 2001-11-14

PRIOR APPLICATION NUMBER: 60/049787

PRIOR FILING DATE: 1997-06-16

PRIOR APPLICATION NUMBER: 60/062250

PRIOR FILING DATE: 1997-10-17

PRIOR APPLICATION NUMBER: 60/065186

PRIOR FILING DATE: 1997-11-12

PRIOR APPLICATION NUMBER: 60/065311

PRIOR FILING DATE: 1997-11-13

PRIOR APPLICATION NUMBER: 60/066770

PRIOR FILING DATE: 1997-11-24

PRIOR APPLICATION NUMBER: 60/075945

PRIOR FILING DATE: 1998-02-25

PRIOR APPLICATION NUMBER: 60/078910

PRIOR FILING DATE: 1998-03-20

PRIOR APPLICATION NUMBER: 60/083322

PRIOR FILING DATE: 1998-04-28

PRIOR APPLICATION NUMBER: 60/084600

PRIOR FILING DATE: 1998-05-07

PRIOR APPLICATION NUMBER: 60/087106

PRIOR FILING DATE: 1998-05-28

PRIOR APPLICATION NUMBER: 60/087607

PRIOR FILING DATE: 1998-06-02

PRIOR APPLICATION NUMBER: 60/087609

PRIOR FILING DATE: 1998-06-02

PRIOR APPLICATION NUMBER: 60/087759

PRIOR FILING DATE: 1998-06-02

PRIOR APPLICATION NUMBER: 60/087827

PRIOR FILING DATE: 1998-06-03

PRIOR APPLICATION NUMBER: 60/088021

PRIOR FILING DATE: 1998-06-04

PRIOR APPLICATION NUMBER: 60/088025

PRIOR FILING DATE: 1998-06-04

PRIOR APPLICATION NUMBER: 60/088026

PRIOR FILING DATE: 1998-06-04

PRIOR APPLICATION NUMBER: 60/088028

PRIOR FILING DATE: 1998-06-04

PRIOR APPLICATION NUMBER: 60/088029

PRIOR FILING DATE: 1998-06-04

PRIOR APPLICATION NUMBER: 60/088030

PRIOR FILING DATE: 1998-06-04

PRIOR APPLICATION NUMBER: 60/088033

PRIOR FILING DATE: 1998-06-04

PRIOR APPLICATION NUMBER: 60/088326

PRIOR FILING DATE: 1998-06-04

PRIOR APPLICATION NUMBER: 60/088167

PRIOR FILING DATE: 1998-06-05

PRIOR APPLICATION NUMBER: 60/088202

PRIOR FILING DATE: 1998-06-05

PRIOR APPLICATION NUMBER: 60/088212

PRIOR FILING DATE: 1998-06-05

PRIOR APPLICATION NUMBER: 60/088217

PRIOR FILING DATE: 1998-06-05

PRIOR APPLICATION NUMBER: 60/088655

PRIOR FILING DATE: 1998-06-09

PRIOR APPLICATION NUMBER: 60/088734

PRIOR FILING DATE: 1998-06-10

PRIOR APPLICATION NUMBER: 60/088738

PRIOR FILING DATE: 1998-06-10

PRIOR APPLICATION NUMBER: 60/088742

PRIOR FILING DATE: 1998-06-10

PRIOR APPLICATION NUMBER: 60/088810

PRIOR FILING DATE: 1998-06-10

PRIOR APPLICATION NUMBER: 60/088824

PRIOR FILING DATE: 1998-06-10

PRIOR APPLICATION NUMBER: 60/088826

PRIOR FILING DATE: 1998-06-10

PRIOR APPLICATION NUMBER: 60/088858

PRIOR FILING DATE: 1998-06-11

PRIOR APPLICATION NUMBER: 60/088861

PRIOR FILING DATE: 1998-06-11

PRIOR APPLICATION NUMBER: 60/088876

PRIOR FILING DATE: 1998-06-11

PRIOR APPLICATION NUMBER: 60/089105

PRIOR FILING DATE: 1998-06-12

PRIOR APPLICATION NUMBER: 60/089440

PRIOR FILING DATE: 1998-06-16

PRIOR APPLICATION NUMBER: 60/089512

PRIOR FILING DATE: 1998-06-16

PRIOR APPLICATION NUMBER: 60/089514

PRIOR FILING DATE: 1998-06-16

PRIOR APPLICATION NUMBER: 60/089532

PRIOR FILING DATE: 1998-06-17

PRIOR APPLICATION NUMBER: 60/089538

PRIOR FILING DATE: 1998-06-17

PRIOR APPLICATION NUMBER: 60/089598

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 30

PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089599  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089600  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089653  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089801  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/089907  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/089908  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/089947  
PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: 60/089948  
PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: 60/089952  
PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: 60/090246  
PRIOR FILING DATE: 1998-06-22  
PRIOR APPLICATION NUMBER: 60/090252  
PRIOR FILING DATE: 1998-06-22  
PRIOR APPLICATION NUMBER: 60/090254  
PRIOR FILING DATE: 1998-06-22  
PRIOR APPLICATION NUMBER: 60/090349  
PRIOR FILING DATE: 1998-06-23  
PRIOR APPLICATION NUMBER: 60/090355  
PRIOR FILING DATE: 1998-06-23  
PRIOR APPLICATION NUMBER: 60/090429  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090431  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090435  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090444  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090445  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090472  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090535  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090540  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090542  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090557  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090676  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090678  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090690  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090694  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090695  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090696  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090862  
PRIOR FILING DATE: 1998-06-26  
PRIOR APPLICATION NUMBER: 60/090863  
PRIOR FILING DATE: 1998-06-26  
PRIOR APPLICATION NUMBER: 60/091360  
PRIOR FILING DATE: 1998-07-01  
PRIOR APPLICATION NUMBER: 60/091478  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091544  
PRIOR FILING DATE: 1998-07-01  
PRIOR APPLICATION NUMBER: 60/091519  
PRIOR FILING DATE: 1998-07-02

PRIOR APPLICATION NUMBER: 60/091626  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091633  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091978  
PRIOR FILING DATE: 1998-07-07  
PRIOR APPLICATION NUMBER: 60/091982  
PRIOR FILING DATE: 1998-07-07  
PRIOR APPLICATION NUMBER: 60/092182  
PRIOR FILING DATE: 1998-07-09

Query Match 4.2%; Score 79; DB 9; Length 867;  
Best Local Similarity 23.8%; Pred. No. 19;  
Matches 59; Conservative 34; Mismatches 85; Indels 70; Gaps 16;

QY 99 TPFNRL-----WYRL--SSRFLSKSYC-----RPFLLYLEAFGLSD 138  
DB 138 TAFPKYINEYNGSYVPGKKEVGLKNSRYNY-TLCRGVYKXGSDYKDY--LTD 194  
QY 139 FLDRQAVYKPELETHSYR-----VSGFVAP-----QYLSLDQRYPIASVWR 185  
DB 195 LITNSV-SFF--RTSKRMTPHBPVLMVSHAPGPDSDAPQY-----SRLFPNAS-QH 245  
QY 186 TLDKNFSLAPDLIDHLGHVPMWLPSPSEFFINMGRLPFKVIEKVALPSKKORIQT 245  
DB 246 ITPSYNAFNPD-----KHWIMRYTPMKPIHME--FTNMLQ-----RKLQTL 287  
QY 246 QSNLIATVRCFWFTVESGLIENHGKAYGAVLISSPOLGHAFIDNVRVLPLEDOIIR 305  
DB 288 MSVDMSMETTYNMLVETGELDN-----TYIVTADHGVIQFGLVXGKSMPEPFD--IR 340  
QY 306 LPFNSTSP 313  
DB 341 VPFYVRGP 348

RESULT 43  
US-09-989-730-84  
Sequence 84, Application US/09989730  
Publication No. US20020197674A1  
GENERAL INFORMATION:  
APPLICANT: Ashkenazi, Avi J.  
APPLICANT: Baker, Kevin P.  
APPLICANT: Botstein, David  
APPLICANT: Deenoyers, Luc  
APPLICANT: Baton, Dan L.  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Fong, Sherman  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gerltsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Goddowski, Paul J.  
APPLICANT: Grimaldi, J. Christopher  
APPLICANT: Gurney, Austin L.  
APPLICANT: Kijavlin, Ivar J.  
APPLICANT: Napier, Mary A.  
APPLICANT: Park, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: P2730PIC69  
CURRENT APPLICATION NUMBER: US/09/989,730  
CURRENT FILING DATE: 2001-11-20  
PRIOR APPLICATION NUMBER: 60/049787  
PRIOR FILING DATE: 1997-06-16  
PRIOR APPLICATION NUMBER: 60/062250



1	PRIOR APPLICATION NUMBER: 60/089512	
2	PRIOR FILING DATE: 1998-06-16	
3	PRIOR APPLICATION NUMBER: 60/089514	
4	PRIOR FILING DATE: 1998-06-16	
5	PRIOR APPLICATION NUMBER: 60/089532	
6	PRIOR FILING DATE: 1998-06-17	
7	PRIOR APPLICATION NUMBER: 60/089538	
8	PRIOR FILING DATE: 1998-06-17	
9	PRIOR APPLICATION NUMBER: 60/089598	
10	PRIOR FILING DATE: 1998-06-17	
11	PRIOR APPLICATION NUMBER: 60/089599	
12	PRIOR FILING DATE: 1998-06-17	
13	PRIOR APPLICATION NUMBER: 60/089600	
14	PRIOR FILING DATE: 1998-06-17	
15	PRIOR APPLICATION NUMBER: 60/089653	
16	PRIOR FILING DATE: 1998-06-17	
17	PRIOR APPLICATION NUMBER: 60/089801	
18	PRIOR FILING DATE: 1998-06-18	
19	PRIOR APPLICATION NUMBER: 60/089907	
20	PRIOR FILING DATE: 1998-06-18	
21	PRIOR APPLICATION NUMBER: 60/089908	
22	PRIOR FILING DATE: 1998-06-18	
23	PRIOR APPLICATION NUMBER: 60/089947	
24	PRIOR FILING DATE: 1998-06-19	
25	PRIOR APPLICATION NUMBER: 60/089948	
26	PRIOR FILING DATE: 1998-06-19	
27	PRIOR APPLICATION NUMBER: 60/089952	
28	PRIOR FILING DATE: 1998-06-19	
29	PRIOR APPLICATION NUMBER: 60/090246	
30	PRIOR FILING DATE: 1998-06-22	
31	PRIOR APPLICATION NUMBER: 60/090252	
32	PRIOR FILING DATE: 1998-06-22	
33	PRIOR APPLICATION NUMBER: 60/090254	
34	PRIOR FILING DATE: 1998-06-22	
35	PRIOR APPLICATION NUMBER: 60/090349	
36	PRIOR FILING DATE: 1998-06-23	
37	PRIOR APPLICATION NUMBER: 60/090355	
38	PRIOR FILING DATE: 1998-06-23	
39	PRIOR APPLICATION NUMBER: 60/090429	
40	PRIOR FILING DATE: 1998-06-24	
41	PRIOR APPLICATION NUMBER: 60/090431	
42	PRIOR FILING DATE: 1998-06-24	
43	PRIOR APPLICATION NUMBER: 60/090435	
44	PRIOR FILING DATE: 1998-06-24	
45	PRIOR APPLICATION NUMBER: 60/090444	
46	PRIOR FILING DATE: 1998-06-24	
47	PRIOR APPLICATION NUMBER: 60/090445	
48	PRIOR FILING DATE: 1998-06-24	
49	PRIOR APPLICATION NUMBER: 60/090472	
50	PRIOR FILING DATE: 1998-06-24	
51	PRIOR APPLICATION NUMBER: 60/090535	
52	PRIOR FILING DATE: 1998-06-24	
53	PRIOR APPLICATION NUMBER: 60/090540	
54	PRIOR FILING DATE: 1998-06-24	
55	PRIOR APPLICATION NUMBER: 60/090542	
56	PRIOR FILING DATE: 1998-06-24	
57	PRIOR APPLICATION NUMBER: 60/090557	
58	PRIOR FILING DATE: 1998-06-24	
59	PRIOR APPLICATION NUMBER: 60/090676	
60	PRIOR FILING DATE: 1998-06-25	
61	PRIOR APPLICATION NUMBER: 60/090678	
62	PRIOR FILING DATE: 1998-06-25	
63	PRIOR APPLICATION NUMBER: 60/090690	
64	PRIOR FILING DATE: 1998-06-25	
65	PRIOR APPLICATION NUMBER: 60/090694	
66	PRIOR FILING DATE: 1998-06-25	
67	PRIOR APPLICATION NUMBER: 60/090695	
68	PRIOR FILING DATE: 1998-06-25	
69	PRIOR APPLICATION NUMBER: 60/090696	
70	PRIOR FILING DATE: 1998-06-25	
71	PRIOR APPLICATION NUMBER: 60/090862	
72	PRIOR FILING DATE: 1998-06-26	
73	PRIOR APPLICATION NUMBER: 60/090863	

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 32

```

; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

```

```

Query Match          4.2%; Score 79; DB 9; Length 867;
Best Local Similarity 23.8%; Pred. No. 19;
Matches 59; Conservative 34; Mismatches 85; Indels 70; Gaps 16;

```

```

QY 99 TPFNNL-----WYLL-SSRFSLKSYC-----PPFLDYLAFGLLSD 138
DB 138 TAFGKYMEVNGSYVPPQKEMWVLLKNSFFNY-FLCNQVKEHGSDYKDY--LTD 194
QY 139 FLHQAVIKFPELETHFSYV-----VSGVAPH-----QYLSLDQRYFPFASVWR 185
DB 195 LITNDGV-SFF--RTSKKMYRPRVPLMTISAAPHGSDSAPQI-----SLRPFNKS-QH 245
QY 186 TLDKDNFSJLPDLIHDLGHVPLHPSSEFFLNKRLFTTKYIEKVALPSKKORIQT 245
DB 246 ITPSYVAPNPD-----KXWIKRYGPKPKPIHME--FTNNLQ-----RKLQTL 287
QY 246 QSNLIAIVRCFWFVESGLIENHEGRKAYGAVLSSPQELGHAFLDNVLEPLDQIIR 305
DB 288 MSYDDMETIYMLVETGELDN-----YIYYTADHGYNHIGQGLVKGKSMYERD--IR 340
QY 306 LPENTSTP 313
DB 341 VPFYVRCF 348

```

```

RESULT 44
US-09-990-436-84
; Sequence 84, Application US/09990436
; Publication No. US20020198148A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Baton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Geider, Hanspeter
; APPLICANT: Gerltsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.

```

```

; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2730P1C14
; CURRENT APPLICATION NUMBER: US/09/990,436
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088217
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088655
; PRIOR FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/088734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088824
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088826
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858

```

us-09-438-185a-1047.rapb

Fri Jan 10 10:56:11 2003

PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/088861  
PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/088876  
PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/089105  
PRIOR FILING DATE: 1998-06-12  
PRIOR APPLICATION NUMBER: 60/089440  
PRIOR FILING DATE: 1998-06-16  
PRIOR APPLICATION NUMBER: 60/089512  
PRIOR FILING DATE: 1998-06-16  
PRIOR APPLICATION NUMBER: 60/089514  
PRIOR FILING DATE: 1998-06-16  
PRIOR APPLICATION NUMBER: 60/089532  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089538  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089598  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089599  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089600  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089653  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089801  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/089907  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/089908  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/089947  
PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: 60/089948  
PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: 60/089952  
PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: 60/090246  
PRIOR FILING DATE: 1998-06-22  
PRIOR APPLICATION NUMBER: 60/090252  
PRIOR FILING DATE: 1998-06-22  
PRIOR APPLICATION NUMBER: 60/090254  
PRIOR FILING DATE: 1998-06-22  
PRIOR APPLICATION NUMBER: 60/090349  
PRIOR FILING DATE: 1998-06-23  
PRIOR APPLICATION NUMBER: 60/090355  
PRIOR FILING DATE: 1998-06-23  
PRIOR APPLICATION NUMBER: 60/090429  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090431  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090435  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090444  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090445  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090472  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090535  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090540  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090542  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090557  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090676  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090678  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090690  
PRIOR FILING DATE: 1998-06-25

PRIOR APPLICATION NUMBER: 60/090694  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090695  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090696  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090862  
PRIOR FILING DATE: 1998-06-26  
PRIOR APPLICATION NUMBER: 60/090863  
PRIOR FILING DATE: 1998-06-26  
PRIOR APPLICATION NUMBER: 60/091360  
PRIOR FILING DATE: 1998-07-01  
PRIOR APPLICATION NUMBER: 60/091478  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091544  
PRIOR FILING DATE: 1998-07-01  
PRIOR APPLICATION NUMBER: 60/091519  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091626  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091633  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091978  
PRIOR FILING DATE: 1998-07-07  
PRIOR APPLICATION NUMBER: 60/091982  
PRIOR FILING DATE: 1998-07-07  
PRIOR APPLICATION NUMBER: 60/092182  
PRIOR FILING DATE: 1998-07-09

Query Match 4.2%; Score 79; DB 9; Length 867;  
Best Local Similarity 23.8%; Pred. No. 19;  
Matches 59; Conservative 34; Mismatches 85; Indels 70; Gaps 15;

QY 99 TPFNNRL-----WYRL--SSRFSLMKSYC-----PRFFLDYLEAFGLSD 138  
DB 138 TAFCKYLNVEYSGVPPGKWKVGLLKNRFRNY--TLCRNGVKRKGSDYSKY--LTD 194  
QY 139 FLDCOAVIKFFETHEFSYYP-----VSGFVAPH-----QYLSLLQDRYPFIASVMR 185  
DB 195 LITNSV-SFP--RTSKMYPRPVLWVISHAAPHGSDSAPOY-----SRLFPNAS-QH 245  
QY 186 TLDRKNSLITDLHLGHVPLHSPFSEFFINMGRFLTKEVKQALPSKKRIOTL 245  
DB 246 IFTSINAPND-----KWMRYTGMKPIHME--FTNMLQ-----RKRIOTL 287  
QY 246 QSNLIAVRCFWFTVESGLIENHEGRKAYGAVLISSPOELGHAFIDNVRLPLELQDIR 305  
DB 288 MSVDSMETIYNMLVEIGELDN-----TYIVTADHYHIGQFGLVKGKSNPYEFD--IR 340  
QY 306 LPFNTSTP 313  
DB 341 VPFYVRGP 348

RESULT 45  
US-09-991-181-84  
; Sequence 84, Application US/09991181  
; Publication No. US20020197615A1  
; GENERAL INFORMATION:  
; APPLICANT: Ashkenazi, Avi J.  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Fong, Sherman  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, J. Christopher  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Kljavin, Ivar J.

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 34

APPLICANT: Napier, Mary A.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tunes, Daniel  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: P2730PIC53  
CURRENT APPLICATION NUMBER: US/09/991,181  
PRIOR FILING DATE: 2001-11-16  
PRIOR APPLICATION NUMBER: 60/049787  
PRIOR FILING DATE: 1997-06-16  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/065186  
PRIOR FILING DATE: 1997-11-12  
PRIOR APPLICATION NUMBER: 60/065311  
PRIOR FILING DATE: 1997-11-13  
PRIOR APPLICATION NUMBER: 60/066770  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/075945  
PRIOR FILING DATE: 1998-02-25  
PRIOR APPLICATION NUMBER: 60/078910  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/083322  
PRIOR FILING DATE: 1998-04-28  
PRIOR APPLICATION NUMBER: 60/084600  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/087106  
PRIOR FILING DATE: 1998-05-28  
PRIOR APPLICATION NUMBER: 60/087607  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/087609  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/087759  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/087827  
PRIOR FILING DATE: 1998-06-03  
PRIOR APPLICATION NUMBER: 60/088021  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088025  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088026  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088028  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088029  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088030  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088033  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088326  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088167  
PRIOR FILING DATE: 1998-06-05  
PRIOR APPLICATION NUMBER: 60/088202  
PRIOR FILING DATE: 1998-06-05  
PRIOR APPLICATION NUMBER: 60/088212  
PRIOR FILING DATE: 1998-06-05  
PRIOR APPLICATION NUMBER: 60/088217  
PRIOR FILING DATE: 1998-06-05  
PRIOR APPLICATION NUMBER: 60/088655  
PRIOR FILING DATE: 1998-06-09  
PRIOR APPLICATION NUMBER: 60/088734  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088738  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088742  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088810  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088824  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088826  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088858  
PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/088861  
PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/088876  
PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/089105  
PRIOR FILING DATE: 1998-06-12  
PRIOR APPLICATION NUMBER: 60/089440  
PRIOR FILING DATE: 1998-06-16  
PRIOR APPLICATION NUMBER: 60/089512  
PRIOR FILING DATE: 1998-06-16  
PRIOR APPLICATION NUMBER: 60/089514  
PRIOR FILING DATE: 1998-06-16  
PRIOR APPLICATION NUMBER: 60/089532  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089538  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089598  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089599  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089600  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089653  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089801  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/089907  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/089908  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/089947  
PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: 60/089948  
PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: 60/089952  
PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: 60/090246  
PRIOR FILING DATE: 1998-06-22  
PRIOR APPLICATION NUMBER: 60/090252  
PRIOR FILING DATE: 1998-06-22  
PRIOR APPLICATION NUMBER: 60/090254  
PRIOR FILING DATE: 1998-06-22  
PRIOR APPLICATION NUMBER: 60/090349  
PRIOR FILING DATE: 1998-06-23  
PRIOR APPLICATION NUMBER: 60/090355  
PRIOR FILING DATE: 1998-06-23  
PRIOR APPLICATION NUMBER: 60/090429  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090431  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090435  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090444  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090445  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090472  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090535  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090540  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090542

us-09-438-185a-1047.rapb

Fri Jan 10 10:56:11 2003

```

PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090676
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090678
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090690
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090694
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090695
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090696
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090862
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/090863
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/091360
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091478
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091544
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091519
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091626
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091633
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091978
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/092182
PRIOR FILING DATE: 1998-07-09

Query Match 4.28; Score 79; DB 9; Length 867;
Best Local Similarity 23.8%; Pred. No. 19;
Matches 59; Conservative 34; Mismatches 85; Indels 70; Gaps 16;

QY 99 TPFNRNL-----WYRLI-SRFSWLSYC-----PRFFLDYLEAFGLSD 138
DB 138 TAFGKVLNEYSVPPCKWEVGLLKNSRFYNY-TLCRNGVKCKHGSYKDY--LTD 194
QY 139 FLHQAVIKFEELTHFSYYP-----VSGFVAPH-----QYLSLLQDRYFFIASVMR 185
DB 195 LINDSV-SFF--RTSKWYHPFVLVWISHMAFGHEDSAQY-----SRLFFNAS-QH 245
QY 186 TLKDNFSLTPDLIHLGHVPLWLLHPSFEFFINMGRFTVKVIEKVALPSKQRIQTL 245
DB 246 ITPSYNAPNPD-----KWIWRYTGPMPKPHME--FTNMLQ-----RKLQTL 287
QY 246 QSNLIAIVRCFTWTVESGLIENHEGRKAYGAVLSSPOELGHAFIDNVVLPLELDQIIR 305
DB 288 MSVDSDMETYNMLNVLGTGLDN-----TIIVTADHGHYHIGQGLVKGSKMEYFD--IR 340
QY 306 LPFNTSTP 313
DB 341 VPFYVRGP 348

RESULT 46
US-09-993-687-84
; Sequence 84, Application US/09993687
; Publication No. US20020198149A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Borstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2730PIC11
; CURRENT APPLICATION NUMBER: US/09/993,687
; PRIOR FILING DATE: 2002-11-14
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212

```

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 36

PRIOR FILING DATE: 1998-06-05  
PRIOR APPLICATION NUMBER: 60/088217  
PRIOR FILING DATE: 1998-06-05  
PRIOR APPLICATION NUMBER: 60/088655  
PRIOR FILING DATE: 1998-06-09  
PRIOR APPLICATION NUMBER: 60/088734  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088738  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088742  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088810  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088824  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088826  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088858  
PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/088861  
PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/088876  
PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/089105  
PRIOR FILING DATE: 1998-06-12  
PRIOR APPLICATION NUMBER: 60/089440  
PRIOR FILING DATE: 1998-06-16  
PRIOR APPLICATION NUMBER: 60/089512  
PRIOR FILING DATE: 1998-06-16  
PRIOR APPLICATION NUMBER: 60/089514  
PRIOR FILING DATE: 1998-06-16  
PRIOR APPLICATION NUMBER: 60/089532  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089538  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089598  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089599  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089600  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089653  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089801  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/089907  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/089908  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/089947  
PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: 60/089948  
PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: 60/089952  
PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: 60/090246  
PRIOR FILING DATE: 1998-06-22  
PRIOR APPLICATION NUMBER: 60/090252  
PRIOR FILING DATE: 1998-06-22  
PRIOR APPLICATION NUMBER: 60/090254  
PRIOR FILING DATE: 1998-06-22  
PRIOR APPLICATION NUMBER: 60/090349  
PRIOR FILING DATE: 1998-06-23  
PRIOR APPLICATION NUMBER: 60/090355  
PRIOR FILING DATE: 1998-06-23  
PRIOR APPLICATION NUMBER: 60/090429  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090431  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090435  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090444  
PRIOR FILING DATE: 1998-06-24

PRIOR APPLICATION NUMBER: 60/090445  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090472  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090535  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090540  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090542  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090557  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090676  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090678  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090690  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090694  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090695  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090696  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090862  
PRIOR FILING DATE: 1998-06-26  
PRIOR APPLICATION NUMBER: 60/090863  
PRIOR FILING DATE: 1998-06-26  
PRIOR APPLICATION NUMBER: 60/091360  
PRIOR FILING DATE: 1998-07-01  
PRIOR APPLICATION NUMBER: 60/091478  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091544  
PRIOR FILING DATE: 1998-07-01  
PRIOR APPLICATION NUMBER: 60/091519  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091626  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091633  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091978  
PRIOR FILING DATE: 1998-07-07  
PRIOR APPLICATION NUMBER: 60/091982  
PRIOR FILING DATE: 1998-07-07  
PRIOR APPLICATION NUMBER: 60/092182  
PRIOR FILING DATE: 1998-07-09

Query Match 4.2% Score 79; DB 9; Length 867;  
Best Local Similarity 23.8% Pred. No. 19;  
Matches 59; Conservative 34; Mismatches 85; Indels 70; Gaps 16;

QY 99 TPFNNRL-----WRL--SSRPSLMKSYC-----PRPFDYLEAGFLSD 138  
DB 138 TAFGKYLNEYNGSVYPCMKEMVGLKNSSFFVYV-TLCNGVKEKGSYSKDY--LTD 194  
QY 139 FLDAQVYKFELETHSYYP-----VSGFVAPH-----OYSLDORYPPIASVNR 185  
DB 195 LITWDSV-SFF--RTSKKMYPHRPVIMWYISHAAPHGEDSAPQY-----SRLEFPNAS-OH 245  
QY 186 TLDKDNSTLPDDLIHDLGHVPMWLPJFSSFFINMGRLEFTKYIEKVALPSKORIQTL 245  
DB 246 ITPEYVAAPND-----KHMIRYTGPKKPIHME--FTNNLQ-----RKSLQTL 287  
QY 246 QSNLIATVRCFWFTVESGLIENHEBKAYGAVILSSPOELGHAFIDNVRLPLELQGITR 305  
DB 288 MSVDSMETHYNNMLVETGELDN-----YYIYTTADHGCHTIGQGLVKGKMPLEFD--IR 340  
QY 306 LPFTSTP 313  
DB 341 VPFYVGP 348  
RESULT 47

us-09-438-185a-1047.rapb

Fri Jan 10 10:56:11 2003

US-09-989-734-84  
 , Sequence 84, Application US/09989734  
 , Publication No. US20030003531A1  
 , GENERAL INFORMATION:  
 , APPLICANT: Ashkenazi, Avi J.  
 , APPLICANT: Baker, Kevin P.  
 , APPLICANT: Botstein, David  
 , APPLICANT: Desnoyers, Luc  
 , APPLICANT: Eaton, Dan L.  
 , APPLICANT: Ferrara, Napoleone  
 , APPLICANT: Fong, Sherman  
 , APPLICANT: Gerber, Hanspeter  
 , APPLICANT: Gerritsen, Mary E.  
 , APPLICANT: Goddard, Audrey  
 , APPLICANT: Godowski, Paul J.  
 , APPLICANT: Grimaldi, J. Christopher  
 , APPLICANT: Gurney, Austin L.  
 , APPLICANT: Kijavini, Ivar J.  
 , APPLICANT: Napier, Mary A.  
 , APPLICANT: Pan, James  
 , APPLICANT: Paoni, Nicholas F.  
 , APPLICANT: Roy, Margaret Ann  
 , APPLICANT: Stewart, Timothy A.  
 , APPLICANT: Tumas, Daniel  
 , APPLICANT: Watanabe, Colin K.  
 , APPLICANT: Williams, P. Mickey  
 , APPLICANT: Wood, William I.  
 , APPLICANT: Zhang, Zemin  
 , TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
 , FILE REFERENCE: P2730P1C64  
 , CURRENT APPLICATION NUMBER: US/09/989,734  
 , CURRENT FILING DATE: 2001-11-19  
 , PRIOR APPLICATION NUMBER: 60/049787  
 , PRIOR FILING DATE: 1997-06-16  
 , PRIOR APPLICATION NUMBER: 60/062250  
 , PRIOR FILING DATE: 1997-10-17  
 , PRIOR APPLICATION NUMBER: 60/065186  
 , PRIOR FILING DATE: 1997-11-12  
 , PRIOR APPLICATION NUMBER: 60/065311  
 , PRIOR FILING DATE: 1997-11-13  
 , PRIOR APPLICATION NUMBER: 60/066770  
 , PRIOR FILING DATE: 1997-11-24  
 , PRIOR APPLICATION NUMBER: 60/075945  
 , PRIOR FILING DATE: 1998-02-25  
 , PRIOR APPLICATION NUMBER: 60/078910  
 , PRIOR FILING DATE: 1998-03-20  
 , PRIOR APPLICATION NUMBER: 60/083322  
 , PRIOR FILING DATE: 1998-04-28  
 , PRIOR APPLICATION NUMBER: 60/084600  
 , PRIOR FILING DATE: 1998-05-07  
 , PRIOR APPLICATION NUMBER: 60/087106  
 , PRIOR FILING DATE: 1998-05-28  
 , PRIOR APPLICATION NUMBER: 60/087607  
 , PRIOR FILING DATE: 1998-06-02  
 , PRIOR APPLICATION NUMBER: 60/087609  
 , PRIOR FILING DATE: 1998-06-02  
 , PRIOR APPLICATION NUMBER: 60/087759  
 , PRIOR FILING DATE: 1998-06-02  
 , PRIOR APPLICATION NUMBER: 60/087827  
 , PRIOR FILING DATE: 1998-06-03  
 , PRIOR APPLICATION NUMBER: 60/088021  
 , PRIOR FILING DATE: 1998-06-04  
 , PRIOR APPLICATION NUMBER: 60/088025  
 , PRIOR FILING DATE: 1998-06-04  
 , PRIOR APPLICATION NUMBER: 60/088026  
 , PRIOR FILING DATE: 1998-06-04  
 , PRIOR APPLICATION NUMBER: 60/088028  
 , PRIOR FILING DATE: 1998-06-04  
 , PRIOR APPLICATION NUMBER: 60/088029  
 , PRIOR FILING DATE: 1998-06-04  
 , PRIOR APPLICATION NUMBER: 60/088030  
 , PRIOR FILING DATE: 1998-06-04  
 , PRIOR APPLICATION NUMBER: 60/088033  
 , PRIOR FILING DATE: 1998-06-04  
 , PRIOR APPLICATION NUMBER: 60/088326  
 , PRIOR FILING DATE: 1998-06-04  
 , PRIOR APPLICATION NUMBER: 60/088167  
 , PRIOR FILING DATE: 1998-06-05  
 , PRIOR APPLICATION NUMBER: 60/088202  
 , PRIOR FILING DATE: 1998-06-05  
 , PRIOR APPLICATION NUMBER: 60/088212  
 , PRIOR FILING DATE: 1998-06-05  
 , PRIOR APPLICATION NUMBER: 60/088217  
 , PRIOR FILING DATE: 1998-06-05  
 , PRIOR APPLICATION NUMBER: 60/088655  
 , PRIOR FILING DATE: 1998-06-09  
 , PRIOR APPLICATION NUMBER: 60/088734  
 , PRIOR FILING DATE: 1998-06-10  
 , PRIOR APPLICATION NUMBER: 60/088738  
 , PRIOR FILING DATE: 1998-06-10  
 , PRIOR APPLICATION NUMBER: 60/088742  
 , PRIOR FILING DATE: 1998-06-10  
 , PRIOR APPLICATION NUMBER: 60/088810  
 , PRIOR FILING DATE: 1998-06-10  
 , PRIOR APPLICATION NUMBER: 60/088824  
 , PRIOR FILING DATE: 1998-06-10  
 , PRIOR APPLICATION NUMBER: 60/088826  
 , PRIOR FILING DATE: 1998-06-10  
 , PRIOR APPLICATION NUMBER: 60/088858  
 , PRIOR FILING DATE: 1998-06-11  
 , PRIOR APPLICATION NUMBER: 60/088861  
 , PRIOR FILING DATE: 1998-06-11  
 , PRIOR APPLICATION NUMBER: 60/088876  
 , PRIOR FILING DATE: 1998-06-11  
 , PRIOR APPLICATION NUMBER: 60/089105  
 , PRIOR FILING DATE: 1998-06-12  
 , PRIOR APPLICATION NUMBER: 60/089440  
 , PRIOR FILING DATE: 1998-06-16  
 , PRIOR APPLICATION NUMBER: 60/089512  
 , PRIOR FILING DATE: 1998-06-16  
 , PRIOR APPLICATION NUMBER: 60/089514  
 , PRIOR FILING DATE: 1998-06-16  
 , PRIOR APPLICATION NUMBER: 60/089532  
 , PRIOR FILING DATE: 1998-06-17  
 , PRIOR APPLICATION NUMBER: 60/089538  
 , PRIOR FILING DATE: 1998-06-17  
 , PRIOR APPLICATION NUMBER: 60/089598  
 , PRIOR FILING DATE: 1998-06-17  
 , PRIOR APPLICATION NUMBER: 60/089599  
 , PRIOR FILING DATE: 1998-06-17  
 , PRIOR APPLICATION NUMBER: 60/089600  
 , PRIOR FILING DATE: 1998-06-17  
 , PRIOR APPLICATION NUMBER: 60/089653  
 , PRIOR FILING DATE: 1998-06-17  
 , PRIOR APPLICATION NUMBER: 60/089801  
 , PRIOR FILING DATE: 1998-06-18  
 , PRIOR APPLICATION NUMBER: 60/089907  
 , PRIOR FILING DATE: 1998-06-18  
 , PRIOR APPLICATION NUMBER: 60/089908  
 , PRIOR FILING DATE: 1998-06-18  
 , PRIOR APPLICATION NUMBER: 60/089947  
 , PRIOR FILING DATE: 1998-06-19  
 , PRIOR APPLICATION NUMBER: 60/089948  
 , PRIOR FILING DATE: 1998-06-19  
 , PRIOR APPLICATION NUMBER: 60/089952  
 , PRIOR FILING DATE: 1998-06-19  
 , PRIOR APPLICATION NUMBER: 60/090246  
 , PRIOR FILING DATE: 1998-06-22  
 , PRIOR APPLICATION NUMBER: 60/090252  
 , PRIOR FILING DATE: 1998-06-22  
 , PRIOR APPLICATION NUMBER: 60/090254  
 , PRIOR FILING DATE: 1998-06-22  
 , PRIOR APPLICATION NUMBER: 60/090349  
 , PRIOR FILING DATE: 1998-06-23  
 , PRIOR APPLICATION NUMBER: 60/090355

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 38

PRIOR FILING DATE: 1998-06-23  
PRIOR APPLICATION NUMBER: 60/090429  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090431  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090435  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090444  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090445  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090472  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090535  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090540  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090542  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090557  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090676  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090678  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090690  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090694  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090695  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090696  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090862  
PRIOR FILING DATE: 1998-06-26  
PRIOR APPLICATION NUMBER: 60/090863  
PRIOR FILING DATE: 1998-06-26  
PRIOR APPLICATION NUMBER: 60/091360  
PRIOR FILING DATE: 1998-07-01  
PRIOR APPLICATION NUMBER: 60/091478  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091544  
PRIOR FILING DATE: 1998-07-01  
PRIOR APPLICATION NUMBER: 60/091519  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091626  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091633  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091978  
PRIOR FILING DATE: 1998-07-07  
PRIOR APPLICATION NUMBER: 60/091982  
PRIOR FILING DATE: 1998-07-07  
PRIOR APPLICATION NUMBER: 60/092182  
PRIOR FILING DATE: 1998-07-09

Query Match 4.2%; Score 79; DB 9; Length 867;  
Best Local Similarity 23.8%; Pred. No. 19;  
Matches 59; Conservative 34; Mismatches 85; Indels 70; Gaps 16;

QY 99 TPFNNL-----WYLL-SSRFSLMKSYC-----PRPFDYEAFLSLD 138  
DB 138 TAFGKLYNEYSYVPBGKEMVLLKNSRFYNY-FLCNGYKEKGSYSKDY--LTD 194  
QY 139 FLHQAVIKFELEHTFSYR-----VSGVAPH-----OYLSLDRIYPLASVWR 185  
DB 195 LITNDVY-SFF--RISKMYRPRVLMYISHAAPGPDASAPOT-----SRLEPNAS-OH 245  
QY 186 TLQDNESLTPLIDHLLGHVWMLHPSFSEFFINMGRLPTKVIEKVOALPSKKORIOTL 245  
DB 246 ITDSYNAAPPD-----KWMIRYTGPMKPIHME--FTNNMLQ-----RKLQTL 287  
QY 246 QSNLIAVRCFMFTVESGLIENHEGRKAYGAVLISSPOELGHAIFIDNVRVLPLELDQIR 305

DB 288 MSVDSMETHIYNMLVETGELDN-----TYIYVTDHGVHIGQFGLVYKSKMPYEPD--IR 340  
QY 306 LEPMTSTP 313  
DB 341 VFFYVRGP 348

RESULT 48  
US-10-028-072-20  
Sequence 20, Application US/10028072  
Publication No. US20030004311A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: DeForge, Laura  
APPLICANT: Desnoyers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerdtzen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang  
TITLE OR INVENTION:  
FILE REFERENCE:  
CURRENT APPLICATION NUMBER: US/10/028.072  
CURRENT FILING DATE: 2001-12-19  
PRIOR APPLICATION NUMBER: 60/049911  
PRIOR FILING DATE: 1997-06-18  
PRIOR APPLICATION NUMBER: 60/056974  
PRIOR FILING DATE: 1997-08-26  
PRIOR APPLICATION NUMBER: 60/059113  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059115  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059117  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059122  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059184  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059263  
PRIOR FILING DATE: 1997-09-18  
PRIOR APPLICATION NUMBER: 60/059352  
PRIOR FILING DATE: 1997-09-19  
PRIOR APPLICATION NUMBER: 60/059588  
PRIOR FILING DATE: 1997-09-19  
PRIOR APPLICATION NUMBER: 60/059836  
PRIOR FILING DATE: 1997-09-24  
PRIOR APPLICATION NUMBER: 60/062250  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/062285  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/062287  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/062814  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/062816  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/063045  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/063082  
PRIOR FILING DATE: 1997-10-31  
PRIOR APPLICATION NUMBER: 60/063127  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/063327



us-09-438-185a-1047.rapb

Fri Jan 10 10:56:11 2003

PRIOR FILING DATE: 1997-10-27  
PRIOR APPLICATION NUMBER: 60/063329  
PRIOR FILING DATE: 1997-10-27  
PRIOR APPLICATION NUMBER: 60/063550  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063561  
PRIOR FILING DATE: 1997-10-28  
PRIOR APPLICATION NUMBER: 60/063704  
PRIOR FILING DATE: 1997-10-29  
PRIOR APPLICATION NUMBER: 60/063733  
PRIOR FILING DATE: 1997-10-29  
PRIOR APPLICATION NUMBER: 60/063735  
PRIOR FILING DATE: 1997-10-29  
PRIOR APPLICATION NUMBER: 60/063738  
PRIOR FILING DATE: 1997-10-29  
PRIOR APPLICATION NUMBER: 60/063755  
PRIOR FILING DATE: 1997-10-17  
PRIOR APPLICATION NUMBER: 60/064248  
PRIOR FILING DATE: 1997-11-03  
PRIOR APPLICATION NUMBER: 60/064809  
PRIOR FILING DATE: 1997-11-07  
PRIOR APPLICATION NUMBER: 60/065186  
PRIOR FILING DATE: 1997-11-12  
PRIOR APPLICATION NUMBER: 60/065846  
PRIOR FILING DATE: 1997-11-17  
PRIOR APPLICATION NUMBER: 60/066364  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: 60/066453  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/066511  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/066770  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/069212  
PRIOR FILING DATE: 1997-12-11  
PRIOR APPLICATION NUMBER: 60/069278  
PRIOR FILING DATE: 1997-12-11  
PRIOR APPLICATION NUMBER: 60/069334  
PRIOR FILING DATE: 1997-12-11  
PRIOR APPLICATION NUMBER: 60/069694  
PRIOR FILING DATE: 1997-12-16  
PRIOR APPLICATION NUMBER: 60/072320  
PRIOR FILING DATE: 1998-01-23  
PRIOR APPLICATION NUMBER: 60/073612  
PRIOR FILING DATE: 1998-02-04  
PRIOR APPLICATION NUMBER: 60/074086  
PRIOR FILING DATE: 1998-02-09  
PRIOR APPLICATION NUMBER: 60/074092  
PRIOR FILING DATE: 1998-02-09  
PRIOR APPLICATION NUMBER: 60/077791  
PRIOR FILING DATE: 1998-03-12  
PRIOR APPLICATION NUMBER: 60/078910  
PRIOR FILING DATE: 1998-03-20  
PRIOR APPLICATION NUMBER: 60/079294  
PRIOR FILING DATE: 1998-03-25  
PRIOR APPLICATION NUMBER: 60/079663  
PRIOR FILING DATE: 1998-02-27  
PRIOR APPLICATION NUMBER: 60/079728  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/080165  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: 60/081203  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081229  
PRIOR FILING DATE: 1998-04-09  
PRIOR APPLICATION NUMBER: 60/081695  
PRIOR FILING DATE: 1998-04-14  
PRIOR APPLICATION NUMBER: 60/081817  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/081818  
PRIOR FILING DATE: 1998-04-15  
PRIOR APPLICATION NUMBER: 60/082999  
PRIOR FILING DATE: 1998-04-24

PRIOR APPLICATION NUMBER: 60/083322  
PRIOR FILING DATE: 1998-04-28  
PRIOR APPLICATION NUMBER: 60/083545  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/084600  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084627  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/084637  
PRIOR FILING DATE: 1998-05-07  
PRIOR APPLICATION NUMBER: 60/085149  
PRIOR FILING DATE: 1998-05-12  
PRIOR APPLICATION NUMBER: 60/085323  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085338  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085339  
PRIOR FILING DATE: 1998-05-13  
PRIOR APPLICATION NUMBER: 60/085579  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085697  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/085704  
PRIOR FILING DATE: 1998-05-15  
PRIOR APPLICATION NUMBER: 60/086414  
PRIOR FILING DATE: 1998-05-22  
PRIOR APPLICATION NUMBER: 60/086430  
PRIOR FILING DATE: 1998-05-22  
PRIOR APPLICATION NUMBER: 60/087106  
PRIOR FILING DATE: 1998-05-28  
PRIOR APPLICATION NUMBER: 60/088026  
PRIOR FILING DATE: 1998-06-04  
PRIOR APPLICATION NUMBER: 60/088730  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088741  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088810  
PRIOR FILING DATE: 1998-06-10  
PRIOR APPLICATION NUMBER: 60/088858  
PRIOR FILING DATE: 19/98-06-11  
PRIOR APPLICATION NUMBER: 60/089532  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089599  
PRIOR FILING DATE: 1998-06-17  
PRIOR APPLICATION NUMBER: 60/089907  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/089947  
PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: 60/090349  
PRIOR FILING DATE: 1998-06-23  
PRIOR APPLICATION NUMBER: 60/090429  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090445  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090538  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090863  
PRIOR FILING DATE: 1998-06-26  
PRIOR APPLICATION NUMBER: 60/091360  
PRIOR FILING DATE: 1998-07-01  
PRIOR APPLICATION NUMBER: 60/091519  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091982  
PRIOR FILING DATE: 1998-07-07

Query Match 4.2%; Score 79; DB 9; Length 867;  
Best Local Similarity 23.8%; Pred.No. 19;  
Matches 59; Conservative 34; Mismatches 85; Indels 70; Gaps 16;

QY 99 TFFNRNL-----WYRL-SRFSLMKSYC-----PRFOLDYLEAFGLSD 138  
Db 138 TAFKGLNEYNSYVPPGWKEVGLLKNSRFYNY-TLCRNGVKEKSGSDYSKDY--LTD 194

Fri Jan 10 10:56:11 2003.

us-09-438-185a-1047.rapb

Page 40

```
OY 139 FLHQAVIKFELETHFSYYP-----VSGFVAPH-----OYLSLDORRYPINASVWR 185
Db 195 LITMDSV-SFF--RTSKMYPRHPLVMI SHAAPHGPEDSAPOY-----SRLEPNNAS-QH 245
OY 186 TLUDKONSLPDLIHLDLGHVWMLHPSESEFIMNGRLFKYVIEKQALPSKKQRIOTL 245
Db 246 ITPSYNAHPD-----KXIMKYTGPKKPIHME--FINMLQ-----KKSLQTL 287
OY 246 OSNLIAVRCFMTFVESGLIEHNGRKAYGAVLISSPOELGHAFIDNVRVLPLELDQIR 305
Db 288 MSYDSDMETIYNMLVETGELDN-----TYIYVTDHGXYHIGQGLVKGKSMPEYFD--IR 340
OY 306 LPFNTSTP 313
Db 341 VPFYVGRP 348

RESULT 49
US-09-989-722-84
Sequence 84: Application us/09989722
Patent No. US20020072067A1
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi J.
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnovers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gertlisen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Auecin L.
APPLICANT: Kijavini, Ivar J.
APPLICANT: Nadier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas P.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Thomas, Daniel
APPLICANT: Williams, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OR INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730P1C63
CURRENT APPLICATION NUMBER: US/09/989,722
PRIOR FILING DATE: 2001-11-19
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02

PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
```

PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/089908  
PRIOR FILING DATE: 1998-06-16  
PRIOR APPLICATION NUMBER: 60/089947  
PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: 60/089948  
PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: 60/089952  
PRIOR FILING DATE: 1998-06-19  
PRIOR APPLICATION NUMBER: 60/090246  
PRIOR FILING DATE: 1998-06-22  
PRIOR APPLICATION NUMBER: 60/090252  
PRIOR FILING DATE: 1998-06-22  
PRIOR APPLICATION NUMBER: 60/090254  
PRIOR FILING DATE: 1998-06-22  
PRIOR APPLICATION NUMBER: 60/090349  
PRIOR FILING DATE: 1998-06-23  
PRIOR APPLICATION NUMBER: 60/090355  
PRIOR FILING DATE: 1998-06-23  
PRIOR APPLICATION NUMBER: 60/090429  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090431  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090435  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090444  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090445  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090472  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090535  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090540  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090542  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090557  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090676  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090678  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090690  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090694  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090695  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/090696  
PRIOR FILING DATE: 1998-06-26  
PRIOR APPLICATION NUMBER: 60/090862  
PRIOR FILING DATE: 1998-06-26  
PRIOR APPLICATION NUMBER: 60/090863  
PRIOR FILING DATE: 1998-06-26  
PRIOR APPLICATION NUMBER: 60/091360  
PRIOR FILING DATE: 1998-07-01  
PRIOR APPLICATION NUMBER: 60/091478  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091544  
PRIOR FILING DATE: 1998-07-01  
PRIOR APPLICATION NUMBER: 60/091519  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091626  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091633  
PRIOR FILING DATE: 1998-07-02  
PRIOR APPLICATION NUMBER: 60/091978  
PRIOR FILING DATE: 1998-07-07  
PRIOR APPLICATION NUMBER: 60/091982  
PRIOR FILING DATE: 1998-07-07  
PRIOR APPLICATION NUMBER: 60/092182  
PRIOR FILING DATE: 1998-07-09

Query Match 4.2%; Score 79; DB 10; Length 867;  
Best Local Similarity 23.8%; Pred. No. 19;  
Matches 59; Conservative 34; Mismatches 85; Indels 70; Gaps 16;

QY 99 TPFENRNL-----WYRL--SSRFLMWSYC-----PRFFLDYLEAFGLLSD 138  
DB 138 TAFEGKYLNEYNGSVVPPGKWEVGLLKNSRFYNY-TLCRNGVKGKSGSDYSKYD--LTD 194  
QY 139 FLDHQAVIKFPELETHFSYYP-----VSGFVAPH-----QYLSLQDRYFFPIASVWR 185  
DB 195 LIITNSV-SFF--RTSKMYRHPVLMVISHAAPHGPDSPAFQY-----SRLFNPNAS-QH 245  
QY 186 TLDDKDNFSLTPDLIHDLGHVFWLLHPSFSEFFINMGRFTVKVIEKVAQALPESKKORIOTL 245  
DB 246 ITFSYVAPNPD-----KHWIMRVTPGPKPIHME--FTNMLQ-----RKRLQTL 287  
QY 246 QSNLIAIVRCFWFTVESGLIENHGRKAYGAVLISSPQELGHAFIDNVRVLPLELDQIIR 305  
DB 288 MSVDDSNMETIYNMLVETGELDN-----TYIVYTADHGYHIGQFLVKGKSMPEYFD--IR 340  
QY 306 LPFNTSTP 313  
DB 341 VFFYVRGP 348

RESULT 50  
US-09-989-723-84  
; Sequence 84, Application US/09989723  
; Patent No. US20020072092A1  
; GENERAL INFORMATION:  
; APPLICANT: Ashkenazi, Avi J.  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Botstein, David  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Ferrara, Napoleone  
; APPLICANT: Forgi, Sherman  
; APPLICANT: Gerber, Hanspeter  
; APPLICANT: Gerttsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, J. Christopher  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Kljavin, Ivar J.  
; APPLICANT: Napier, Mary A.  
; APPLICANT: Pan, James  
; APPLICANT: Paoni, Nicholas F.  
; APPLICANT: Roy, Margaret Ann  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tunas, Daniel  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Williams, P. Mickey  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
; TITLE OF INVENTION: Acids Encoding the Same  
; FILE REFERENCE: P2730P1662  
; CURRENT APPLICATION NUMBER: US/09/989, 723  
; CURRENT FILING DATE: 2001-11-19  
; PRIOR APPLICATION NUMBER: 60/049787  
; PRIOR FILING DATE: 1997-06-16  
; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/065186  
; PRIOR FILING DATE: 1997-11-12  
; PRIOR APPLICATION NUMBER: 60/065311  
; PRIOR FILING DATE: 1997-11-13  
; PRIOR APPLICATION NUMBER: 60/066770  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/075945  
; PRIOR FILING DATE: 1998-02-25  
; PRIOR APPLICATION NUMBER: 60/078910

[illegible]

us-09-438-185a-1047.rapb

Fri Jan 10 10:56:11 2003

; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091633  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: 60/091978  
; PRIOR FILING DATE: 1998-07-07  
; PRIOR APPLICATION NUMBER: 60/091982  
; PRIOR FILING DATE: 1998-07-07  
; PRIOR APPLICATION NUMBER: 60/092182  
; PRIOR FILING DATE: 1998-07-09

Query Match 4.2%; Score 79; DB 10; Length 867;  
Best Local Similarity 23.8%; Pred. No. 19;  
Matches 59; Conservative 34; Mismatches 85; Indels 70; Gaps 16;  
QY 99 TFFNRL-----WVRL--SRFSLWKSVC-----PRFFLDYLEAFGLLSD 138  
Db 138 TAFGKYLNEYNGSVPPGKWKVGLLKNSRFVNY-TLCRNGVKEKHGSDYSKYD--LTD 194  
QY 139 FLDHOAVIKFESELETHFSYYP-----VSGFVAPH-----QYLSLLQDRYFPPIASVMR 185  
Db 195 LTINDSV-SFP--RTSKMYFHRPVLVMIASHAAPHCPEDSAPQY-----SRLFPNAS-QH 245  
QY 186 TLDKONFSLTDPDLIHDLLGHVPMLLHPSFSEFPINMGRFTKVIKQALPSKKORIOTL 245  
Db 246 ITFSYNAFND-----KHWIMRYTGMKFIHME--FTNMLQ-----RKRLQTL 287  
QY 246 QSNLIAVRCFWFTVESGLTENHEGRKAYGAVLISSPQELGHAFIDNVRVLPLELDQIIR 305  
Db 288 MSYDDSMETIYNMLVETGELDN-----TYIVYTADGHYHIGQFGLVKGKSMPEFD--IR 340  
QY 306 LPFNTSTP 313  
Db 341 VPEYVRGP 348

Search completed: January 9, 2003, 17:13:45  
Job time : 52 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2003, 14:46:38 ; Search time 26 Seconds  
(without alignments)  
409.658 Million cell updates/sec

Title: US-09-438-185A-1047

Perfect score: 1989  
Sequence: 1 VHYCERTLDPKYLKIALKL.....ESIPLYNQEKYLSGFEVLQ 362

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pap.\*

2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pap.\*

3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pap.\*

4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pap.\*

5: /cgn2\_6/ptodata/1/iaa/PCTUS.COMB.pap.\*

6: /cgn2\_6/ptodata/1/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	223	11.8	498	1	US-08-009-075-5
2	220	11.6	497	1	US-08-009-075-4
3	98.5	5.2	1024	2	US-09-091-117-5
4	95	5.0	1226	4	US-08-540-804-12
5	95	5.0	1226	2	US-08-218-265-12
6	95	5.0	1226	3	US-08-521-872-12
7	95	5.0	1226	4	US-08-590-399-12
8	92	4.9	4536	4	US-09-180-422B-27
9	91.5	4.8	872	3	US-08-766-014-2
10	91	4.8	535	3	US-08-369-822C-24
11	91	4.8	535	3	US-08-582-776C-39
12	91	4.8	535	3	US-08-434-831B-36
13	91	4.8	2183	2	US-08-348-891A-7
14	91	4.8	2183	2	US-08-905-817-7
15	90.5	4.8	615	1	US-08-484-106-6
16	90.5	4.8	615	1	US-08-484-106-6
17	90	4.8	625	4	US-09-134-001C-4504
18	90	4.8	654	4	US-09-134-001C-3261
19	89	4.7	392	3	US-08-911-853-35
20	89	4.7	392	4	US-09-479-409-35
21	89	4.7	392	4	US-09-479-453-35
22	87.5	4.6	309	3	US-08-996-338-22
23	87.5	4.6	521	3	US-08-996-338-20
24	87.5	4.6	541	1	US-08-604-333-2
25	87.5	4.6	541	3	US-09-110-618-2
26	87.5	4.6	541	4	US-09-173-151A-28
27	87.5	4.6	541	4	US-09-578-178-2

28	86.5	4.6	990	2	US-08-392-625-20	Sequence 20, Appl
29	86.5	4.6	990	2	US-08-466-961A-20	Sequence 20, Appl
30	86.5	4.6	990	2	US-08-645-193B-15	Sequence 15, Appl
31	86	4.6	656	4	US-09-134-001C-4322	Sequence 4322, Ap
32	84.5	4.5	360	4	US-09-810-347-4	Sequence 4, Appl
33	84.5	4.5	372	4	US-09-810-347-2	Sequence 2, Appl
34	83.5	4.4	1365	6	5194600-4	Patent No. 5194600
35	83	4.4	491	2	US-08-942-819-2	Sequence 2, Appl
36	82	4.3	284	4	US-08-913-159-2	Sequence 2, Appl
37	82	4.3	861	3	US-09-022-875-2	Sequence 2, Appl
38	81.5	4.3	307	4	US-09-810-347-5	Sequence 5, Appl
39	81.5	4.3	608	1	US-08-766-014-3	Sequence 3, Appl
40	81.5	4.3	1708	1	US-08-493-092-2	Sequence 2, Appl
41	81.5	4.3	1708	1	US-08-508-836A-2	Sequence 2, Appl
42	81.5	4.3	3056	2	US-08-508-836A-8	Sequence 8, Appl
43	81.5	4.3	3056	2	US-08-629-001A-3	Sequence 3, Appl
44	81.5	4.3	3056	2	US-08-874-266-2	Sequence 2, Appl
45	81.5	4.3	3056	4	US-08-642-274D-3	Sequence 3, Appl
46	81.5	4.3	3056	4	US-08-952-127D-3	Sequence 3, Appl
47	81.5	4.3	3056	4	US-08-952-014C-3	Sequence 3, Appl
48	81.5	4.3	3057	4	US-09-360-416-3	Sequence 3, Appl
49	81	4.3	323	2	US-08-540-804-18	Sequence 18, Appl
50	81	4.3	323	3	US-08-218-265-18	Sequence 18, Appl
51	81	4.3	323	3	US-08-521-872-18	Sequence 18, Appl
52	81	4.3	323	4	US-08-590-399-18	Sequence 18, Appl
53	81	4.3	474	4	US-08-961-083-60	Sequence 60, Appl
54	81	4.3	1118	3	US-08-724-354D-2	Sequence 2, Appl
55	81	4.3	1118	3	US-09-270-984A-2	Sequence 2, Appl
56	80.5	4.3	377	2	US-07-863-169A-1	Sequence 1, Appl
57	80.5	4.3	377	2	US-08-429-964-1	Sequence 1, Appl
58	80.5	4.3	377	3	US-07-935-087-1	Sequence 1, Appl
59	80.5	4.3	377	5	PCT-US93-08062-1	Sequence 1, Appl
60	80.5	4.3	1365	4	US-09-376-330-18	Sequence 18, Appl
61	79.5	4.2	787	1	US-08-574-763-2	Sequence 2, Appl
62	79.5	4.2	981	2	US-08-649-046-2	Sequence 2, Appl
63	79	4.2	207	3	US-08-996-338-23	Sequence 23, Appl
64	79	4.2	674	1	US-08-803-973-7	Sequence 7, Appl
65	79	4.2	674	1	US-08-803-972-7	Sequence 7, Appl
66	79	4.2	707	1	US-08-803-973-12	Sequence 12, Appl
67	79	4.2	707	1	US-08-803-972-12	Sequence 12, Appl
68	79	4.2	1456	1	US-08-803-973-2	Sequence 2, Appl
69	79	4.2	1456	1	US-08-803-972-2	Sequence 2, Appl
70	79	4.2	3418	2	US-08-639-501-2	Sequence 2, Appl
71	79	4.2	3418	2	US-08-603-753D-4	Sequence 4, Appl
72	79	4.2	3418	3	US-09-044-946-2	Sequence 2, Appl
73	79	4.2	3418	3	US-08-755-587-44	Sequence 44, Appl
74	79	4.2	3418	3	US-09-044-908-2	Sequence 2, Appl
75	79	4.2	3418	4	US-09-099-753-4	Sequence 4, Appl
76	79	4.2	3418	4	US-08-986-106-4	Sequence 4, Appl
77	79	4.2	3457	2	US-08-416-603-4	Sequence 4, Appl
78	78.5	4.2	796	4	US-08-737-109-31	Sequence 31, Appl
79	78.5	4.2	3056	4	US-09-360-416-2	Sequence 2, Appl
80	78	4.1	317	4	US-09-327-984A-37	Sequence 37, Appl
81	78	4.1	670	4	US-09-134-001C-2940	Sequence 2940, Ap
82	78	4.1	3066	4	US-08-953-137-12	Sequence 12, Appl
83	77.5	4.1	514	4	US-09-413-814-82	Sequence 82, Appl
84	77.5	4.1	514	4	US-09-413-814-94	Sequence 94, Appl
85	77.5	4.1	577	4	US-09-315-794-32	Sequence 32, Appl
86	77.5	4.1	577	4	US-09-389-341-32	Sequence 32, Appl
87	77.5	4.1	739	4	US-09-134-001C-3586	Sequence 3586, Ap
88	77.5	4.1	905	4	US-09-134-001C-3782	Sequence 3782, Ap
89	77.5	4.1	971	2	US-08-724-354D-22	Sequence 22, Appl
90	77.5	4.1	971	3	US-09-270-984A-22	Sequence 22, Appl
91	77.5	4.1	971	3	US-09-177-431-8	Sequence 8, Appl
92	77.5	4.1	1089	3	US-08-375-300-2	Sequence 2, Appl
93	77.5	4.1	1089	3	PCT-US95-16930-2	Sequence 2, Appl
94	77.5	4.1	1089	5	PCT-US95-16930-2	Sequence 2, Appl
95	77	4.1	284	2	US-08-424-641B-2	Sequence 2, Appl
96	77	4.1	284	2	US-08-826-439-2	Sequence 2, Appl
97	77	4.1	284	2	US-08-826-439-2	Sequence 2, Appl
98	77	4.1	578	1	US-08-766-014-4	Sequence 4, Appl
99	77	4.1	645	1	US-07-779-172A-3	Sequence 3, Appl
100	77	4.1	749	1	US-08-369-796-8	Sequence 8, Appl

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rai

Page 2

101 77 4.1 749 2 US-08-852-091-8 Sequence 8, Appl1  
102 77 4.1 749 2 US-08-820-754-8 Sequence 8, Appl1  
103 77 4.1 749 3 US-08-856-652-8 Sequence 8, Appl1  
104 77 4.1 749 3 US-08-856-869-8 Sequence 7, Appl1  
105 77 4.1 749 3 US-08-012-710-7 Sequence 8, Appl1  
106 77 4.1 749 3 US-08-848-547-8 Sequence 8, Appl1  
107 77 4.1 749 4 US-08-586-273-7 Sequence 8, Appl1  
108 77 4.1 749 4 US-08-956-653A-8 Sequence 8, Appl1  
109 77 4.1 749 5 PCT-US95-17025-8 Sequence 2, Appl1  
110 77 4.1 844 5 US-09-564-805-227 Sequence 2, Appl1  
111 77 4.1 1663 2 US-08-793-126-1 Sequence 1, Appl1  
112 77 4.1 1663 4 US-09-132-771-1 Sequence 22, Appl1  
113 77 4.1 1663 4 US-09-142-334-22 Sequence 4, Appl1  
114 76.5 4.0 311 2 US-08-602-359A-41 Sequence 12, Appl1  
115 76.5 4.0 614 1 US-08-295-814B-12 Sequence 8, Appl1  
116 76.5 4.0 614 1 US-08-291-299-8 Sequence 12, Appl1  
117 76.5 4.0 614 4 US-09-343-361-12 Sequence 8, Appl1  
118 76.5 4.0 614 5 PCT-US95-10579-8 Sequence 2, Appl1  
119 76.5 4.0 723 4 US-09-134-001C-5060 Sequence 2, Appl1  
120 76.5 4.0 1865 1 US-08-588-985-2 Sequence 2, Appl1  
121 76.5 4.0 246 4 US-09-134-001C-5292 Sequence 71, Appl1  
122 76 4.0 349 1 US-08-118-270-71 Sequence 84, Appl1  
123 76 4.0 349 5 PCT-US93-08528-71 Sequence 95, Appl1  
124 76 4.0 604 1 US-08-224-657-84 Sequence 95, Appl1  
125 76 4.0 604 1 US-08-224-657-95 Sequence 95, Appl1  
126 76 4.0 604 1 US-08-354-138-84 Sequence 95, Appl1  
127 76 4.0 604 4 US-03-354-138-95 Sequence 95, Appl1  
128 76 4.0 604 4 US-08-724-354D-4 Sequence 95, Appl1  
129 76 4.0 604 4 US-08-124-354D-4 Sequence 95, Appl1  
130 76 4.0 1043 2 US-08-270-010-3 Sequence 3, Appl1  
131 76 4.0 2254 2 US-08-790-519-3 Sequence 3, Appl1  
132 76 4.0 2254 2 US-08-134-001C-3703 Sequence 130, Appl1  
133 76 4.0 496 4 US-08-757-653-190 Sequence 71, Appl1  
134 75.5 4.0 833 2 US-08-823-516-71 Sequence 132, Appl1  
135 75.5 4.0 833 3 US-08-758-314-132 Sequence 132, Appl1  
136 75.5 4.0 833 3 US-08-758-314-132 Sequence 132, Appl1  
137 75.5 4.0 833 3 US-08-758-314-132 Sequence 132, Appl1  
138 75.5 4.0 833 3 US-08-758-314-132 Sequence 132, Appl1  
139 75.5 4.0 833 3 US-08-758-314-132 Sequence 132, Appl1  
140 75.5 4.0 833 3 US-08-758-314-132 Sequence 132, Appl1  
141 75 4.0 888 4 US-09-134-001C-3032 Sequence 8, Appl1  
142 75 4.0 564 3 US-08-425-843-8 Sequence 8, Appl1  
143 75 4.0 1257 1 US-08-049-783-2 Sequence 6, Appl1  
144 75 4.0 1257 1 US-08-158-232-6 Sequence 6, Appl1  
145 75 4.0 1257 1 US-08-304-626-6 Sequence 6, Appl1  
146 75 4.0 1257 1 US-08-316-301A-6 Sequence 6, Appl1  
147 75 4.0 1257 2 US-08-611-928-6 Sequence 6, Appl1  
148 75 4.0 1257 3 US-09-173-891-6 Sequence 6, Appl1  
149 75 4.0 1257 4 US-09-076-137-6 Sequence 6, Appl1  
150 75 4.0 1257 5 PCT-US92-03624-6 Sequence 6, Appl1

ALIGNMENTS

RESULT 1  
US-08-009-075-5  
Sequence 5, Application US/08009075  
Patent No. 5300436  
GENERAL INFORMATION:  
APPLICANT: GOLDSTEIN, Menek  
APPLICANT: WU, Jing  
APPLICANT: FRIEDHOFF, David  
APPLICANT: FRIEDHOFF, Arnold J.  
TITLE OF INVENTION: GENETICALLY MODIFIED TYROSINE  
TITLE OF INVENTION: HYDROXYLASE AND USES THEREOF  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.

ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/08/009,075  
APPLICATION NUMBER: US/08/009,075  
FILING DATE: 1993/01/26  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: TOWNSEND, GUY K.  
REGISTRATION NUMBER: 34,033  
REFERENCE/DOCKET NUMBER: GOLDSTEIN-1A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 498 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-009-075-5  
Query Match 11.8%; Score 223; DB 1; Length 498;  
Best Local Similarity 26.7%; Pred. No. 1,4e-15;  
Matches 65; Conservative 39; Mismatches 103; Indels 36; Gaps 5;  
112 SRRSLKLS-----YCPRFLLDYLAFFGLSDFDLH-----QAVIKFELETHF 155  
Db AEIATKVEYVTLKELVATHACREHLGFOLEIRYCYGRBDSI PQLIEDVSRLKERTGF 286  
Qy 156 SYVPSGFVAHQYLSLDODRYFPFASVMTLDKDNFSLTPDILHDLGHVPLHPSPS 215  
Db 287 QLRVPAQLLSARDFLSLAFRVFOCTOYIRHASSPMHSPEDDCHELGHVPLMDRTFA 346  
Qy 216 EFFINMRLLTKYIEKVALPSKORIOTLOSNIATVRCFWFTVESGLIENHGRKAYG 275  
Db 347 QFSODIG-----LASIGASDEIRKLSI-----VYMFVFEGLCKQNGELKAYG 390  
Qy 276 AVLISSPOELGHFIDNVRVLPLELDQIIRLPFNTSTPOETLFIRIF---DELVELTS 331  
Db 391 AGLISSYGLHLSLSEPEVRAPFDPTAAGPYDDQYQPYVFSSEFNDADKLNAYAS 450  
Qy 332 KLE 334  
Db 451 RIQ 453  
RESULT 2  
US-08-009-075-4  
Sequence 4, Application US/08009075  
Patent No. 5300436  
GENERAL INFORMATION:  
APPLICANT: GOLDSTEIN, Menek  
APPLICANT: WU, Jing  
APPLICANT: FRIEDHOFF, David  
APPLICANT: FRIEDHOFF, Arnold J.  
TITLE OF INVENTION: GENETICALLY MODIFIED TYROSINE  
TITLE OF INVENTION: HYDROXYLASE AND USES THEREOF  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible



```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/009,075
; FILING DATE: 19930126
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: TOWNSEND, GUY K.
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: GOLDSTEIN=1A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO. 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 497 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-009-075-4

Query Match 11.6%; Score 220; DB 1; Length 497;
Best Local Similarity 26.7%; Pred. No. 3e-15;
Matches 65; Conservative 38; Mismatches 104; Indels 36; Gaps 5;

QY 112 SRSFSLMKS-----YCPRFPLDYLEAFGLLSDFLDH-----QAVIKPFLETHF 155
DB 226 ABEIATWKEVVTTLKGLYATHACGEHLEAFALLERFSGYREDNIPQLEDVSRFLKERTGF 285

QY 156 SYYPVSGFVAPHOYLSLODRYPIASVMTLKDKNFSLTDLIDHLLGHVPWLLHPSFS 215
DB 286 QLRPVAGLSARDFLASLAFVQCQYIRHASSPMHSPEDCCHELLGHVPMLADRTFA 345

QY 216 EFTINMGLFTKVEKQVALPSKKORIOTLQSNLAIIVRCFWFTVESGLIENHEGKAYG 275
DB 346 QFSQDIG-----LASGASDEIEKLSLTS-----WFTVEGLCKONGEVKAYG 389

QY 276 AVLTSSQELGHAFIDNVRVLPLELDQIIRLPNTSTPOETLFSIRHF-----DELVELTS 331
DB 390 AGLSSVYGLLHCLSEEPFADPEAAVQYQDQYQSVYVSESFSDAKDKLRSYAS 449

QY 332 KLE 334
DB 450 RIQ 452

```

## RESULT 3

```

US-09-091-117-5
; Sequence 5, Application US/09091117
; Patent No. 6171589
; GENERAL INFORMATION:
; APPLICANT: The University of Melbourne
; TITLE OF INVENTION: Mycoplasma Recombinant Polypeptides and
; TITLE OF INVENTION: Vaccines
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GREENLEE, WINNER and SULLIVAN P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/091,117
; FILING DATE: 12 JUNE 1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU96/00803
; FILING DATE: 13-DEC-1996

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PN7127
; FILING DATE: 13-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: WINNER, Ellen P.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: +1 303 499 8080
; TELEFAX: +1 303 499 8089
; INFORMATION FOR SEQ ID NO. 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1024 amino acids
; TYPE: amino acid
; MOLECULE TYPE: protein
; ORIGINAL SOURCE: Mycoplasma genitalium
; ORGANISM: Mycoplasma genitalium
; US-09-091-117-5

Query Match 5.2%; Score 98.5; DB 4; Length 1024;
Best Local Similarity 19.9%; Pred. No. 0.13;
Matches 68; Conservative 41; Mismatches 134; Indels 99; Gaps 11;

QY 29 QNSOSLQRAYSTPYSYRIILQKENKEK---QALAHKICISILEFFKNLLFVHLLSLSKN 85
DB 564 QQTDSLKNLFSV---IGDILSETNVNKTILHAVKNELLSLVETASTLKIKHL----- 613

QY 86 QREGCSTDMAVVSTPPFNENLWYLLSSRFSLSKSYCPRFPLDYLEAFGLLSDFLDHQA 145
DB 614 -----NVQYKVLVDKFKLSFK-----ELNFFPDTKDITPT 647

QY 146 IK--FFLETH-----PSYYPVSGFVAPHOYLSLODRYPIASVMTLKDKNF 192
DB 648 IKKVLFESENYKTLKKYENEGPGYHMAKIVPGTFNSAENTFYSAI-----DKT 698

QY 193 SLTDLIDHLLGHVPWLLHPSSEPFINMGRFLT-----KVIEKVALPSKKORIOT 244
DB 699 KSIRDLFADML-FGKLSLVNSDSDFIKINGSFTLYKHGDNLLPNYHSLITKNVGYQI 757

QY 245 LOSNLIAIVRCFWFTVESGLIENHEGKAYGAVLISSPOELG----- 286
DB 758 VNVAFHIDARLLTAELONTVFSNPK-----FVIKSPVELSKSLFEVWKTIFENSVNOI 810

QY 287 ----HAFIDNVRVLPLELDQIIRLPNTSTPOETLFSIRHPD 324
DB 811 LKGYTFKONLKFPPFKADGSGRLSPDLKPKDQRPVPPAFVD 852

```

## RESULT 4

```

US-08-540-804-12
; Sequence 12, Application US/08540804
; Patent No. 5919666
; GENERAL INFORMATION:
; APPLICANT: Young, Richard A.
; APPLICANT: Koleske, Anthony J.
; APPLICANT: Thompson, Craig M.
; APPLICANT: Chao, David M.
; TITLE OF INVENTION: No. 5919666el Factors Which Modify Gene
; TITLE OF INVENTION: Transcription and Methods of Use Therefor
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/540,804

```

FILED DATE: 11-OCT-1995  
CLASSIFICATION: 424  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/521,872  
FILING DATE: 21-AUG-1995  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,265  
FILING DATE: 25-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: WH194-03A2  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1226 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-540-804-12

Query Match 5.0%; Score 95; DB 2; Length 1226;  
Best Local Similarity 18.8%; Pred. No. 0.4; Indels 204; Gaps 25;  
Matches 98; Conservative 65; Mismatches 155; Indels 204; Gaps 25;

QY 12 YILKIALKROSLFONSSLOPASTPYRYRII-----LQENKEKQALAHKCI 65  
DB 13 YILE--KLFDMTHNHYNDSQL--RTWKROISYFLKLGNCYSRLINKS--IFHMLV 65  
QY 66 SILEFFKNLFF---VHLSSKNQREGCSTMAVST-----PFF----- 102  
DB 66 EFINKMENFEFLSLHILMIFWNDICQIDTNAFVAATITSSQKEPFLVTXTDMLHK 125  
QY 103 -----NRLMYRLSSRFSL-----WKS 120  
DB 126 YIVSSSKSMINDENYIINDIKKNNKIKLNILKLSLILKIPQOSLEVFIFPTSNMEI 185  
QY 121 YCRPFF-----LDYLEAFGLSDPFDHOAVIKFEBLETHFSYVSGFVAP 166  
DB 186 YKPLFEIVSNADTQNSDMKKLELISYRNESLKNNSISR-----NVIASNN 236  
QY 167 HOYLSLDQRYFPIASV-----RTLDKON-----FSLTPDLIHDLGHVP 207  
DB 237 DFOULTVTCQFPLKSCIQNCIDTQFTKLD--DNPTFDPWPTVVDNPLTHKIIQLIL 295  
QY 208 WLLHPS--FSEFFINNGRLFTKVIKVOALPSKKQRIQTLSNL--IAIVRCFWPTVESG 263  
DB 296 WSHIPSRQPDHYESN-----OLVAKLILL-----RINSTDEDLHEFOLEDAIWSLVFO- 343  
QY 264 LIENHEGRK-----AYGAV-----LISSP-----OELGHAFFID-- 291  
DB 344 LAKNFSAQKRVVSYMPSLYRLNLITTYGIIVKPYIRKLISGGLLYQDSNDKVFVHQ 403  
QY 292 ---NVRVLPLELDQIIRLPFNSTPOETLFSIRHDELYELVETSKL----- 333  
DB 404 LLINEKISPLMKSQYNNMVLRYNMEVDVYKFEIFNFDOLVETIRIKRIISNDITNLQLS 463  
QY 334 -----EMMLD---OGLIESIPLYNQEKLISGFEVLC 361  
DB 464 KTPLSIKIMVAEWYLSHLCGILSV--NRTVLKIKFKIFC 502

RESULT 5  
US-08-218-265-12  
Sequence 12, Application US/08218265  
Patent No. 5922585  
GENERAL INFORMATION:  
APPLICANT: Young, Richard A.  
APPLICANT: Koleske, Anthony J.  
APPLICANT: Thompson, Craig M.  
TITLE OF INVENTION: No. 5922585el Factors which Modify Gene

TITLE OF INVENTION: Transcription and Methods of Use Thereof  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: US  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/218,265  
FILING DATE: 25-MAR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: WH194-03  
TELEPHONE: 617-861-9540  
TELEFAX: 617-861-6240  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1226 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-218-265-12

Query Match 5.0%; Score 95; DB 2; Length 1226;  
Best Local Similarity 18.8%; Pred. No. 0.4;  
Matches 98; Conservative 65; Mismatches 155; Indels 204; Gaps 25;

QY 12 YILKIALKROSLFONSSLOPASTPYRYRII-----LQENKEKQALAHKCI 65  
DB 13 YILE--KLFDMTHNHYNDSQL--RTWKROISYFLKLGNCYSRLINKS--IFHMLV 65  
QY 66 SILEFFKNLFF---VHLSSKNQREGCSTMAVST-----PFF----- 102  
DB 66 EFINKMENFEFLSLHILMIFWNDICQIDTNAFVAATITSSQKEPFLVTXTDMLHK 125  
QY 167 HOYLSLDQRYFPIASV-----RTLDKON-----FSLTPDLIHDLGHVP 207  
DB 237 DFOULTVTCQFPLKSCIQNCIDTQFTKLD--DNPTFDPWPTVVDNPLTHKIIQLIL 295  
QY 208 WLLHPS--FSEFFINNGRLFTKVIKVOALPSKKQRIQTLSNL--IAIVRCFWPTVESG 263  
DB 296 WSHIPSRQPDHYESN-----OLVAKLILL-----RINSTDEDLHEFOLEDAIWSLVFO- 343  
QY 264 LIENHEGRK-----AYGAV-----LISSP-----OELGHAFFID-- 291  
DB 344 LAKNFSAQKRVVSYMPSLYRLNLITTYGIIVKPYIRKLISGGLLYQDSNDKVFVHQ 403  
QY 292 ---NVRVLPLELDQIIRLPFNSTPOETLFSIRHDELYELVETSKL----- 333  
DB 404 LLINEKISPLMKSQYNNMVLRYNMEVDVYKFEIFNFDOLVETIRIKRIISNDITNLQLS 463  
QY 334 -----EMMLD---OGLIESIPLYNQEKLISGFEVLC 361  
DB 464 KTPLSIKIMVAEWYLSHLCGILSV--NRTVLKIKFKIFC 502

RESULT 6  
 US-08-521-872-12  
 ; Sequence 12, Application US/08521872  
 ; Patent No. 6015682  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Young, Richard A.  
 ; APPLICANT: Koleske, Anthony J.  
 ; APPLICANT: Thompson, Craig M.  
 ; APPLICANT: Chao, David M.  
 ; TITLE OF INVENTION: No. 6015682el Factors Which Modify Gene  
 ; TITLE OF INVENTION: Transcription and Methods of Use Therefor  
 ; NUMBER OF SEQUENCES: 37  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
 ; STREET: Two Militia Drive  
 ; CITY: Lexington  
 ; STATE: Massachusetts  
 ; COUNTRY: USA  
 ; ZIP: 02173  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/521,872  
 ; FILING DATE: 31-AUG-1995  
 ; CLASSIFICATION: 436  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/218,265  
 ; FILING DATE: 25-MAR-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Granahan, Patricia  
 ; REGISTRATION NUMBER: 32,227  
 ; REFERENCE/DOCKET NUMBER: WHI94-03A  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617-861-6240  
 ; TELEFAX: 617-861-9540  
 ; INFORMATION FOR SEQ ID NO: 12:  
 ; SEQUENCE CHARACTERISTICS:  
 ; TYPE: amino acid  
 ; LENGTH: 1226 amino acids  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-521-872-12  
 Query Match 5.0%; Score 95; DB 3; Length 1226;  
 Best Local Similarity 18.8%; Pred. No. 0.4;  
 Matches 98; Conservative 65; Mismatches 155; Indels 204; Gaps 25;  
 QY 12 YILKIALKRLQSLFFONSQSLORAYSTPSYRII-----LQENKEQALARKCI 65  
 DB 13 YILE---KLIFDMTHYNDSQL-RTWKQISYFLKLCNYSLLKNE---IFHHLV 65  
 QY 66 SILEFPKNLFP-----VHLLSKNQREGCTDMVVST-----PPF-----102  
 DB 66 EFINKMENFEPLSLHILMIFWINDICQIDTNAFVAATITSSQKEPFLVTKITDMLLHK 125  
 QY 103 -----NRNLWYRLSSRFSL-----WKS 120  
 DB 126 YIIVSSKSMINDENYIINDIKKKNKILKILSLILKIFQSLFVFIPTSNWEI 185  
 QY 121 YCPRF-----LDYLEAFGLSDFLDQAVIKFPELETHFSYYPVSGFVAP 166  
 DB 186 YKELLFEIVSNADTNQNSDMKKLELISYRNESLKNSSIR-----NVIMSASNAN 236  
 QY 167 HQVLSLQDRYEPILASVM-----RTLDKDN-----FSLTPDLIHDLGHVP 207  
 DB 237 DFQITVTCTQPKFLSCQLNCIDTQTCKLDD-DNTEFDWYTDYQDPLTHKIIQL 295  
 QY 208 WLLHPS--FSEFINKMGLFTKVEKQALPSKKQRIQTLQSNL--IAIVRCFWFTVESG 263

DB 296 WSIHPSRQFDHYESN-----QLVAKLLLL-----RINSTDEDLHEFQIEDAINSLVFQ- 343  
 QY 264 LIENHEGRK-----AYGAV-----LISSP-----QELGHAFID-- 291  
 DB 344 LAKNFAQRVSVYMMPSLYRLNLILITYGIIKVPYIRKLISSGLLYLQDSNDKCFVHVQ 403  
 QY 292 ---NVRVLPLELDQIIRLPENTSTPQETLSIRHFDLVELTSKL----- 333  
 DB 404 LLNLKISPLMKSQYNWLVNRNMEYDKFYEFNFQDLVEITEQIKMRLSNDIITNLQLS 463  
 QY 334 -----BWLID---OGLLESIPLYNQEKVLSGFVLC 361  
 DB 464 KTLPSIKIMVAEWYLSHLCSGILSV---NRTVLLKIFKIFC 502  
 RESULT 7  
 US-08-590-399-12  
 ; Sequence 12, Application US/08590399  
 ; Patent No. 6214588  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Young, Richard A.  
 ; APPLICANT: Koleske, Anthony J.  
 ; APPLICANT: Thompson, Craig M.  
 ; APPLICANT: Chao, David M.  
 ; TITLE OF INVENTION: No. 6214588el Factors Which Modify Gene  
 ; TITLE OF INVENTION: Transcription and Methods of Use Therefor  
 ; NUMBER OF SEQUENCES: 39  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
 ; STREET: Two Militia Drive  
 ; CITY: Lexington  
 ; STATE: Massachusetts  
 ; COUNTRY: USA  
 ; ZIP: 02173  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/590,399  
 ; FILING DATE: 26-JAN-1996  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/540,804  
 ; FILING DATE: 11-OCT-1995  
 ; APPLICATION NUMBER: US 08/521,872  
 ; FILING DATE: 31-AUG-1995  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/218,265  
 ; FILING DATE: 25-MAR-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Granahan, Patricia  
 ; REGISTRATION NUMBER: 32,227  
 ; REFERENCE/DOCKET NUMBER: WHI94-03A3  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617-861-6240  
 ; TELEFAX: 617-861-9540  
 ; INFORMATION FOR SEQ ID NO: 12:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1226 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-590-399-12

Query Match 5.0%; Score 95; DB 4; Length 1226;  
 Best Local Similarity 18.8%; Pred. No. 0.4;  
 Matches 98; Conservative 65; Mismatches 155; Indels 204; Gaps 25;  
 QY 12 YILKIALKRLQSLFFONSQSLORAYSTPSYRII-----LQENKEQALARKCI 65

Fri Jan 10 10:56:11 2003

US-09-438-185a-1047.rai

Page 6

Db 13 YILE--KLIPTMTNHYNDSQL--RTWKROQISYFLKLLGNCYSRLINKE---IFPHMLV 65  
Qy 66 SILEFFKNLLF---VHLISLSKNOREGCTDMAVST-----PFF----- 102  
Db 66 EFINKQNEFEFLPLSHLMTLFWNDICOIDNAPVAVITTSQKREFFLVYTKITDMLHK 125  
Qy 103 -----NENLWYRLSSRFL-----WKS 120  
Db 126 YIVSSSKMINDENYIINDIKNNKIKLNTLKLISLLIKIFQEOSLEVFIPTTSNWEI 185  
Qy 121 YCPFFF-----LDYLEAFGLSDFLDHOAVIKFPELETHFSYVSGFVAP 166  
Db 186 YKPLFEIVSNADTQNSDMKKKELLISYRNSLKNNSSIR-----NYIWSASNAN 236  
Qy 167 HOYLSLDRYPFIASVM-----RTLDKDN-----FSLTPDLIHLGLHVP 207  
Db 237 DPOLITVTCQKPKLSCIQNCIDTQFTKLD-DNPTFDPMPFYVDQNPFLTMKIIQLIL 295  
Qy 208 WLLHPS--FSEFIMNGRLFTVIEKVOALPSKKQRIQIOTQSNL--IAIVRCFWFTVESG 263  
Db 296 WSIHPSRQPDHYESN-----QVAKLL-----RINSTDEDLHFOJEDALMSLVFC- 343  
Qy 264 LIENHGRK-----AYGAV-----LISSP-----OELGHAFID-- 291  
Db 344 LAKFSAQKQVVSYYMPSLYRLNLITTYGILKVPYIRKLISGLLYIODSNDKFXVQ 403  
Qy 292 ---NVRVPLELDQIIRLPFNTSTPOETLFSIRHPDELVELTSKL----- 333  
Db 404 LLINKISPLMKSQYNNVYLRNWEYDKFEIENFDQIVETQIKRRLISNDITNLQLS 463  
Qy 334 -----EWMLD---QGLLESIPLYNOEKYLSGEVLC 361  
Db 464 KTLPLSIKIMVAEWYLSHLCGILSSV--NRTVLKIKFKIFC 502

RESULT 8  
US-09-180-422B-27  
Sequence 27, Application US/09180422B  
Patent No. 644644  
GENERAL INFORMATION:  
APPLICANT: BRUCKDORFER, KARL R  
TITLE OF INVENTION: ANTICOAGULANT PEPTIDE FRAGMENTS DERIVED  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHAYE, P.C.  
STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22201  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/180,422B  
FILING DATE: 07-Dec-1998  
CLASSIFICATION: <unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: SADOFF, B.J.  
REGISTRATION NUMBER: 36663  
REFERENCE/DOCKET NUMBER: 117-268  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 7038164000  
TELEFAX: 7038164100  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 27:  
US-09-180-422B-27  
Query Match 4.9%; Score 92; DB 4; Length 4536;  
Best Local Similarity 20.7%; Pred. No. 6.2; Indels 68; Gaps 12;  
Matches 61; Conservative 48; Mismatches 118;  
Qy 3 YCERTLPKYLIKALTKQSLSPFQNSQSLQRAYSTPSYYR--IIQKENEKQAL 59  
Db 4184 YTBRELCTMPFREVGVTLQVYSKYVNGSEIL-----FSYFODLVITLFFR-----L 4230  
Qy 60 ARHKCISILEFFKNLLFVHLISLSKNOREGCTDMAVSTPFFNRNLW-----YELLS 113  
Db 4231 RKHKLIVISGYREL-----LKOLSKBAQEVFKAIQSLKTTEVL--RNLQDLLOFIFOLLIED 4285  
Qy 114 RFLSWKSYCPREFLDYL--EAFGLSDFLDHOAVIKFPELETHFSYVSGFVAP----- 166  
Db 4286 NIKQKMKKFTYILNTIYQEIINTIFNDYIPY--VFKLKENLCLNKHKFEIQLNELQEA 4343  
Qy 167 -----HOYLSLDRYPFIASVMRT-----LDKDNFSLTPDLI----- 199  
Db 4344 SOELQOHQIYIMALREYTPDSIVGWTYKYTELEKIVSLKKNLVALKDFHSEYIVSAS 4403  
Qy 200 ---HDLGHVPWLLHPSFSFEPFIMNGRLFTKYIEKVOALPSKKQRIQIOTQSNLIA 251  
Db 4404 NFTSQLSQVEQFLHRNIQIYLSILTPDQKKEKIEIABLSYTAQEI--IKSQALH 4456

RESULT 9  
US-08-766-014-2  
Sequence 2, Application US/08766014  
Patent No. 5744312  
GENERAL INFORMATION:  
APPLICANT: Mamone, Joseph A.  
APPLICANT: Davis, Maria  
APPLICANT: Sha, Dan  
TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASE  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
STREET: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/766,014  
FILING DATE: Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below:  
APPLICATION NUMBER: 60/008,688  
FILING DATE: December 15, 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 222/104  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 469-1600  
TELEFAX: (213) 955-0440  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

LENGTH: 872 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-08-766-014-2

Query Match 4.8%; Score 91.5; DB 1; Length 872;  
 Best Local Similarity 17.5%; Pred. No. 0.57;  
 Matches 74; Conservative 70; Mismatches 141; Indels 139; Gaps 19;

QY 14 LKIALKRLQSLSFFQNSQSORAYSTPYRILOKEXKQALARKHCISILEFFKN 73  
 DB 224 IKENIENKELAIMSKRLATIKRDIPIEIDFEYKVKFNEEK-----LLELFNK 273  
 QY 74 LLFVHLLSLSKNQREGCGTDMVAVSTPFNNRLMYRLISGRFSLMKSYCPRFLLDYLEAF 133  
 DB 274 LEFFSLDNKKE-----SSIEIVD-----NHKVEKWSK-----VDIKELV 309  
 QY 134 GLSDFLDHOAVIKFPELETHPSYYPV-----SGFVAPHOYLSLLQRIYPIASV 183  
 DB 310 TLQD-----NNIAFYLLIVEGEEKIAPFGKOTVIDVQTE-----DL 351  
 QY 184 MRLDKONFSLTPDLIHDLI-----GHVPMLLHPSFSEFFINMGRFLT 226  
 DB 352 KEIPEKEDFEETHEIKDFLVRLSYKGIKESKYIDTAVMAYLLNPSSENY--DLDRVLK 409  
 QY 227 KVIKQALPS-----KKQRIOTLOSNIIA-----IVRCFW----- 257  
 DB 410 KYL-KVD-VPSYEGIFGKGRDKKIEIDENILADYICSCVYLFDLKEKLMNFIEEMDM 467  
 QY 258 -----FTVSGHJENHGRKAYGAVLISGP-QELGHAFIDNVRLPLBLDQIR--LPPN 309  
 DB 468 KXLLLEIEMPLIVEVLKSMESVSGFTLDKVELKSELKSIDDRIGEI---LDKIYKEAGYQN 524  
 QY 310 TSTPOE-TLFSIRHFDLVELTSKLEWMLDQGLLESIFLYN-----OEKY 353  
 DB 525 VNSPKQLSEFLFKLNLFPVTKTKGTSTDSVLRLQVLYNDIVSDIIEYRQLTKLKSTY 584  
 QY 354 LSGF 357  
 DB 585 IDGF 588

RESULT 10  
 US-08-369-822C-24  
 Sequence 24: Application US/08369822C  
 Patent No. 6015860

GENERAL INFORMATION:  
 APPLICANT: Lipkin, W. I.  
 APPLICANT: Briese, Thomas  
 APPLICANT: Kliche, Stefanie  
 APPLICANT: Schneider, Patrick A.  
 APPLICANT: Stiltz, Lothar  
 APPLICANT: Schneemann, Anette

TITLE OF INVENTION: Borna Disease Viral Sequences,  
 TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous  
 TITLE OF INVENTION: System Diseases  
 NUMBER OF SEQUENCES: 46  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Fulbright & Jaworski, L.L.P.  
 STREET: 865 South Figueroa Street, 29th Floor  
 CITY: Los Angeles  
 STATE: California  
 COUNTRY: USA  
 ZIP: 90017-2571

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: WINDOWS NT-WORDPERFECT 8.0  
 SOFTWARE: ASCII (DOS) TEXT  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/369,822C  
 FILING DATE: 06-JAN-1995

CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Churchill, Margaret A. (Ph.D.)  
 REGISTRATION NUMBER: 39,944  
 REFERENCE/DOCKET NUMBER: 1279-194XX  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 213/892-9200  
 TELEFAX: 213/890-4518  
 INFORMATION FOR SEQ ID NO. 24:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 535 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 US-08-369-822C-24

Query Match 4.8%; Score 91; DB 3; Length 535;  
 Best Local Similarity 20.6%; Pred. No. 0.31;  
 Matches 78; Conservative 46; Mismatches 119; Indels 136; Gaps 18;

QY 57 QALARHKCISILEFFKNLLFVHLLSLSKNQREGCGTDMVAVSTPFNNRLMYRLISLRFS 116  
 DB 1 BGLTHEQVDNWKSFAGVKF-----GCFMPLSLDSD--LTMYLKDKALAAALQR 46  
 QY 117 LMKSYCPRRFLDYLEAFGLSLDFLDHOAVIKFPELETHPSYYPVSGFVAPHOYLSLLQDR 176  
 DB 47 EMDSVYKPEFLRYDPPKGTGS-----RLVDVFLNDSSFPDYDVIMVVGAY---LHDP 98  
 QY 177 YETIASVMRTLDKDNFSLTPDLIHDLLGHVPMLLHPSFSEFFINMGRFLTGVKVKQALP 236  
 DB 99 EFNL-----SISLQEKIK-----TGELFAKTYKMA-- 127  
 QY 237 SKKQRIOTLOSNIIA-----IVRCFWFTVBSG-----LIENHGRKAY 274  
 DB 128 -----COVIAENLISNGIKYFKDNGMAKQDQTLKALHLAVSGVPMKDKESHG-- 178  
 QY 275 GAVLISSPQELGHAFIDNVRLP--LELDQIRLPFTNSTPOE----- 315  
 DB 179 GPVLKTSRSPVHTSTNRVRAAGFGIFQVIRQDQTDHPENMEAVETVSAFITTLKK 238  
 QY 316 -----TLFSIRHFDLVELTSKLEWMLDQGLL-----ESIPLY-- 348  
 DB 239 YCLNWRVETISLFAQR-LNEIYGLSPFFQWLKRLTSLVLYSDPHCPDLDHAHPLYKV 297  
 QY 349 -NOE---KY-LSGFVFLCQ 362  
 DB 298 PNDQIFIKYPMWGIEGYCQ 316

RESULT 11  
 US-08-582-776C-39  
 Sequence 39: Application US/08582776C  
 Patent No. 6079510

GENERAL INFORMATION:  
 APPLICANT: Lipkin, W. I.  
 APPLICANT: Briese, Thomas  
 APPLICANT: Kliche, Stefanie  
 APPLICANT: Schneider, Patrick A.  
 APPLICANT: Stiltz, Lothar  
 APPLICANT: Schneemann, Anette

TITLE OF INVENTION: Borna Disease Viral Sequences,  
 TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous  
 TITLE OF INVENTION: System Diseases  
 NUMBER OF SEQUENCES: 61  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Fulbright & Jaworski, L.L.P.  
 STREET: 865 South Figueroa Street, 29th Floor  
 CITY: Los Angeles  
 STATE: California  
 COUNTRY: USA  
 ZIP: 90017-2576

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.ra1

Page 8

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: WINDOWS NT
SOFTWARE: ASCII DOS TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/582.776C
FILING DATE: 04-JAN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/369,822
FILING DATE: 06-JAN-1995
APPLICATION NUMBER: US 08/434,831
FILING DATE: 04-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Churchill, Margaret A.
REGISTRATION NUMBER: 39,944
REFERENCE/DOCKET NUMBER: 1279-194C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213/680-4518
TELEFAX: 213/682-9200
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 535 amino acids
TYPE: amino acid
MOLECULE TYPE: peptide
HYPOTHEICAL: NO
ANTI-SENSE: NO
US-08-582-776C-39

Query Match
Best Local Similarity 20.6%, Pred. No. 0.31;
Matches 78; Conservative 46; Mismatches 119; Indels 136; Gaps 18;

QY 57 QALARHKCISILEFFKNLLFVHLISLKNQREGCSTDMAVVSTPFNNRLMWRLLSSRFS 116
D 1 EGLTHEQCVDNWKSFGVGF-----GCFMPLSLDSD--LTMVLKDKALALQR 46
QY 117 LKSYCPREFLDYLEAFGLSDFLDHQAVIKFELETHFSYVSGFVAPHQYLSLDDR 176
D 47 EMDSVYKPEFLRDPKGTGS-----RLVDVFLNDSFPDYIMVVSAGV---LHDP 98
QY 177 YFPIASVMTLDKDNFSLTPDLIHDLGHVWMLHPSFSEFFINMGRLETKYIEKVALP 236
D 99 EFNL-----SISLOEKEIKE-----TGRLEAKMTYKARA-- 127
QY 237 SKQRIQTLQSNLIA-----IVRCFWFTVESG---LIENHEGRKAY 274
D 128 -----CQVIAENLISNGICKYFKDNGMAKDQDLTKALHTLAVSGVKDKLKHSHRG--- 178
QY 275 GAVLISSPOELGHAFIDNVRLP--LELDQIIRLPFNTSTPOE----- 315
D 179 GPVLKTYSRPVTSTRNVRAAKGFIPPOVIRQDODTDHPENMEAYETVSATITTDLKK 238
QY 316 -----TLFSIRHPDELVELTSKLEMM---LDOGLT-----ESILPY-- 348
D 239 YCLNWRYETISLFAQR-LNEIYGLPSFPQWLHKRLKLETSVLVSDPHCPDLDLHPIYKV 297
QY 349 -NOE---KY-LSGFEVLCO 362
D 298 PNDQIFIKYPMGIGGYCO 316

RESULT 12
US-08-434-831B-36
Sequence 36, Application US/08434831B
Patent No. 6113905
GENERAL INFORMATION:
APPLICANT: Lipkin, W. I.
APPLICANT: Briese, Thomas
APPLICANT: Kliche, Stefanie
```

```

APPLICANT: Schneider, Patrick A.
APPLICANT: Stitz, Lothar
APPLICANT: Schneemann, Anette
TITLE OF INVENTION: Borna Disease Viral Sequences,
TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous
TITLE OF INVENTION: System Diseases
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski, L.L.P.
STREET: 865 South Figueroa Street, 29th Floor
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90017-2571
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: WINDOWS NT-WORDPERFECT 8.0
SOFTWARE: ASCII (DOS) TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/434,831B
FILING DATE: 04-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/369,822
FILING DATE: 06-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Churchill, Margaret A. (Ph.D.)
REGISTRATION NUMBER: 39,944
REFERENCE/DOCKET NUMBER: 1279-194C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213/680-4518
TELEFAX: 213/682-9200
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 535 amino acids
TYPE: amino acid
MOLECULE TYPE: peptide
HYPOTHEICAL: NO
ANTI-SENSE: NO
US-08-434-831B-36

Query Match
Best Local Similarity 20.6%, Pred. No. 0.31;
Matches 78; Conservative 46; Mismatches 119; Indels 136; Gaps 18;

QY 57 QALARHKCISILEFFKNLLFVHLISLKNQREGCSTDMAVVSTPFNNRLMWRLLSSRFS 116
D 1 EGLTHEQCVDNWKSFGVGF-----GCFMPLSLDSD--LTMVLKDKALALQR 46
QY 117 LKSYCPREFLDYLEAFGLSDFLDHQAVIKFELETHFSYVSGFVAPHQYLSLDDR 176
D 47 EMDSVYKPEFLRDPKGTGS-----RLVDVFLNDSFPDYIMVVSAGV---LHDP 98
QY 177 YFPIASVMTLDKDNFSLTPDLIHDLGHVWMLHPSFSEFFINMGRLETKYIEKVALP 236
D 99 EFNL-----SISLOEKEIKE-----TGRLEAKMTYKARA-- 127
QY 237 SKQRIQTLQSNLIA-----IVRCFWFTVESG---LIENHEGRKAY 274
D 128 -----CQVIAENLISNGICKYFKDNGMAKDQDLTKALHTLAVSGVKDKLKHSHRG--- 178
QY 275 GAVLISSPOELGHAFIDNVRLP--LELDQIIRLPFNTSTPOE----- 315
D 179 GPVLKTYSRPVTSTRNVRAAKGFIPPOVIRQDODTDHPENMEAYETVSATITTDLKK 238
QY 316 -----TLFSIRHPDELVELTSKLEMM---LDOGLT-----ESILPY-- 348
D 239 YCLNWRYETISLFAQR-LNEIYGLPSFPQWLHKRLKLETSVLVSDPHCPDLDLHPIYKV 297
QY 349 -NOE---KY-LSGFEVLCO 362
D 298 PNDQIFIKYPMGIGGYCO 316
```

DB 298 PNDQIFIKYPMGGIEGYCQ 316

RESULT 13

US-08-348-891A-7

; Sequence 7, Application US/08348891A

; Patent No. 5654136

; GENERAL INFORMATION:

; APPLICANT: SASAKI, Keiko

; APPLICANT: MORI, Takayuki

; TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,

; TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR

; TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: YOUNG & THOMPSON

; STREET: 745 South 23rd Street

; CITY: Arlington

; STATE: Virginia

; COUNTRY: USA

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patent in Release #1.0, Version #1.30

; CURRENT APPLICATION NUMBER: US/08/348,891A

; FILING DATE: 25-NOV-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/848,400

; FILING DATE: 10-MAR-1992

; APPLICATION NUMBER: JP 3-293625

; FILING DATE: 14-OCT-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: PATCH, Andrew J.

; REGISTRATION NUMBER: 32,925

; REFERENCE/DOCKET NUMBER: KP-7501

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-521-2297

; TELEFAX: 703-685-0573

; TELEX: 248425 EMBON

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2183 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-348-891A-7

Query Match 4.8%; Score 91; DB 1; Length 2183;

Best Local Similarity 20.6%; Pred. No. 2.6;

Matches 78; Conservative 46; Mismatches 119; Indels 136; Gaps 18;

QY 57 QALARHKCISILEFFKNLLFVHLLSLSKNORCGCTDMAVSTPFFNRLNMYRLSSRFS 116

DB 429 EGLTHQCCVDNWKSFAGVKF-----GCFMPLSLDS--LTMWLDKALAAALQ 474

QY 117 LWSKSYCFRFLDLEAFGLSLDLDHQAIVKFFFELETHFSYYPVSGFVAPHQYLSLQDR 176

DB 475 EMDSVYPKEFLAYDPKGTGS-----RLVDVFLNDSFDPDYDMVYVSGAY---LHDP 526

QY 177 YPIASVMTLKDKNFTLTPDLIHLLGHVPLLLHFSFSEFFINMGRFLTKVIEKVALP 236

DB 527 EFWL-----SYSLKEKE-----IVRCFWFTVESC-----LIENHEGRKAY 555

QY 237 SKQRTOTLQSNLIA-----IVRCFWFTVESC-----LIENHEGRKAY 274

DB 556 -----CQVTAENLISNGIKGFKNGMAKDEHDLTKALHTLAVSGVPKDLKESHRG----- 606

QY 275 GAVLISSPOBLGHAFIDNVRVLP--LELDQIIRLPNTSTPQE----- 315

DB 607 GPVLKTVSRSPVHTSTRNRAAKGFIGFPQVIRQDQDTPHPENMEAYETVSAFITDLKK 666

QY 316 -----TLFSIRHFDDELVELTSKLEWM---LQGLL-----ESIPLY-- 348

DB 667 YCLNWRVETISLFAQR-LNEIYGLPSFPQMLHRLKRLTSVLVYSDPHCPDLDLDAHPLYKV 725

QY 349 -NOE---KY-LSGFEVLQ 362

DB 726 PNDQIFIKYPMGGIEGYCQ 744

RESULT 14

US-08-905-817-7

; Sequence 7, Application US/08905817

; Patent No. 5824777

; GENERAL INFORMATION:

; APPLICANT: SASAKI, Keiko

; APPLICANT: MORI, Takayuki

; APPLICANT: MAKINO, Satoshi

; TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,

; TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR

; TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: YOUNG & THOMPSON

; STREET: 745 South 23rd Street

; CITY: Arlington

; STATE: Virginia

; COUNTRY: USA

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patent in Release #1.0, Version #1.30

; CURRENT APPLICATION NUMBER: US/08/905,817

; FILING DATE: 04-AUG-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/348,891

; FILING DATE: 25-NOV-1994

; APPLICATION NUMBER: US 07/848,400

; FILING DATE: 10-MAR-1992

; APPLICATION NUMBER: JP 3-293625

; FILING DATE: 14-OCT-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: PATCH, Andrew J.

; REGISTRATION NUMBER: 32,925

; REFERENCE/DOCKET NUMBER: KP-7501A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-521-2297

; TELEFAX: 703-685-0573

; TELEX: 248425 EMBON

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2183 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-905-817-7

Query Match 4.8%; Score 91; DB 2; Length 2183;

Best Local Similarity 20.6%; Pred. No. 2.6;

Matches 78; Conservative 46; Mismatches 119; Indels 136; Gaps 18;

QY 57 QALARHKCISILEFFKNLLFVHLLSLSKNORCGCTDMAVSTPFFNRLNMYRLSSRFS 116

DB 429 EGLTHQCCVDNWKSFAGVKF-----GCFMPLSLDS--LTMWLDKALAAALQ 474

## MOLECULE TYPE: protein

US-08-484-105-6

Query Match 4.8%; Score 90.5; DB 1; Length 615;

Best Local Similarity 20.7%; Pred. No. 0.43; Matches 64; Conservative 54; Mismatches 110; Indels 81; Gaps 16;

QY 117 LKSYCPREFLDYLEAFGLSDFLDHQAVIKFPELETHFSYVSGFVAPHOYLSLLDR 176  
DB 475 EWSVYKPEFLKDYDPKGTGS-----RLVYVFLNDSFPPYIMVYVSGAV---LNDP 526  
QY 177 YPIASVWKTLDKNFSLTPDLIHDLGHWMLHPSFSFPIINMRLFTKYEYQALP 236  
DB 527 EFWL-----SYSLKEKKE-----TRLPKMTYVQRA-- 555  
QY 237 SKKRIQTLQSNLIA-----IVRCFWTVESG---LIENHEGRKAY 274  
DB 556 -----CQVIAENLISNGIGKYFKONGAKDEHDLTKALHTLAVSGVPKDKESHKG--- 606  
QY 275 GAVLISSPOLGHAFIDNVRLP--LELDQIIRLPFTSTPOE-----ESIPLY-- 348  
DB 607 GPVLKTYSRSPVHTSTRVRAAKGFIGFPQYIRQDQDTHDPENMEAYETVSAFITTDLKK 666  
QY 316 -----TLFSIRHPELVELTSKLEMM--LDQGLP----- 666  
DB 667 YCLNMRYETISLFAQR-LNEIYGLSPFQWILHKLSTSVLYVSDPHCPDLDLHILPLKV 725  
QY 349 -NQE--KY-LSGFVLCO 362  
DB 726 PNDQIFIKYPMGIGIERYCO 744

## RESULT 15

US-08-484-105-6  
Sequence 6, Application US/08484105  
Patent No. 5589341

## GENERAL INFORMATION:

APPLICANT: STILLMAN, Bruce  
APPLICANT: BELL, Stephen P  
APPLICANT: KOBAYASHI, Ryuji  
APPLICANT: RINE, Jasper  
APPLICANT: FOSS, Margit  
APPLICANT: MCNALLY, Francis J  
APPLICANT: LAURENSEN, Patricia  
APPLICANT: HERSKOWITZ, Ira  
APPLICANT: LI, Joachim J  
APPLICANT: GAVIN, Kimberly  
TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,105

## FILING DATE:

## CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Osman Ph.D., Richard Aron  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771

## INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:  
LENGTH: 615 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

QY 24 LSLFPONSGLORAYSTPYS-----YRILQENKEKQALARKKCSILEFPNLTLP 76  
DB 336 MSYFONAFSV--FTDPNVVDFLNDDYLTLL-----SKCPTEFMEVGLIK 369  
QY 77 VH-----LSSLNKGEGGS--TMAYVSTP-----FNRNLRLLSSRSLSKSYC 122  
DB 370 QHAPADILSLTLNKRGLGEFFVFLVRENPINGHAKFVARFLBELNTITNELIY- 428  
QY 123 PRFLDYLEAFGLSDFLDHQAVIKFPELETHFSYVPV-----SGFVAPH-- 167  
DB 429 -----HNLIGLDSYLDLRWSACKEYKDRLHFE--PIDTLPQELFTLDRSGLTQSYF 480  
QY 168 -QVSLDQDYFFPIASVWKTLDKNF-SLTPDL--IHDLLGVPMLLHP-----SFSRP 217  
DB 481 PYSKSNIEDNLSEQVLSLDEKNYDTLSGDLKIMAPVIGQLFKLYRANMTIYVDF 540  
QY 218 FIMNGRLFTK--VIEKQALPSKKORIQTLOS-----NLIAIVRCFWTVESGLIEYN 268  
DB 541 YIARETLPEKEITLNFIRDPSTYKLELATPDAPDKVALILFQMAIFAFENMGLIK-F 599  
QY 269 EGRAYGAV 277  
DB 600 QSTKSYDLV 608

## RESULT 16

US-08-484-106-6  
Sequence 6, Application US/08484106  
Patent No. 5614618

## GENERAL INFORMATION:

APPLICANT: STILLMAN, Bruce  
APPLICANT: BELL, Stephen P  
APPLICANT: KOBAYASHI, Ryuji  
APPLICANT: RINE, Jasper  
APPLICANT: FOSS, Margit  
APPLICANT: MCNALLY, Francis J  
APPLICANT: LAURENSEN, Patricia  
APPLICANT: HERSKOWITZ, Ira  
APPLICANT: LI, Joachim J  
APPLICANT: GAVIN, Kimberly  
TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,106

## FILING DATE:

## CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Osman Ph.D., Richard Aron  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771



```
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 615 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-494-106-6

Query Match 4.8%; Score 90.5; DB 1; Length 615;
Best Local Similarity 20.7%; Pred. No. 0.43;
Matches 64; Conservative 54; Mismatches 110; Indels 81; Gaps 16;

QY 24 LSLFQNSQSLQRAYSTPYS-----YYRILOKENKEKOALARKICISILEFFKNLLF 76
DB 326 MSYFQNAFVS---FIDPVNVDPLNDYLLKIL-----SRCTFMFPFVEGLIK 369
QY 77 VH-----LLSLSKNOREGCS---TDMVVVSTP-----FFNRNLWYLLSSRSFSLWKSVC 122
DB 370 QHAPADEILSLTNKNGLEEFFEFLVRENPIGHAKFVARFLEEELNITNENLIELY- 428
QY 123 PRFFLDYLEAFGLSLDPLDHOAVIKFPELETHESYYPV-----SQFVAPH-- 167
DB 429 -----HNLGLKLSYLDWSACKKEYKDLHFE---PIDTFQELFTLDNRSGLLTQSIF 480
QY 168 -OYLSLLQDRYFPFIASVMRTLDKDNF-SLTPDL-----IHDLGHPVMLLHP- -SFGEF 217
DB 481 PSYKNIENLLSWEQVLPSELDKENTIDILSGDLKIMAPVLGQIFKLYREANMTINIDF 540
QY 218 FIMGRLFTK--VIEKQVALPSKKQRIQTLQS-----NLIAIVRCFWFTVESGLIENH 268
DB 541 YIATRETLPEKEILFIRKDPSTNKLLEAETPDAPDKVALILFMQAFAPENNGLIK-F 599
QY 269 EGRKAYGAV 277
DB 600 QSTKSYDLV 608

RESULT 17
US-09-134-001C-4504
; Sequence 4504, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR FILING DATE: 1997-11-08
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4504
; LENGTH: 625
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4504

Query Match 4.8%; Score 90; DB 4; Length 625;
Best Local Similarity 20.5%; Pred. No. 0.5;
Matches 91; Conservative 71; Mismatches 172; Indels 110; Gaps 23;

QY 11 KYILKIALKRLQSLSLFFQNSQSLQRAY-----STPYSYRIILQK- -ENKEQA 58
DB 132 KFDIYLSIKKKQGI-MIDASTTSISNAVLHINQLTDDDFKVENLILQELPOAHRKIKQI 190
QY 59 LARH--KCSILEFFKNLLFVHLLSKN---QREGCSTD-----MAVSTPFPFNRLW 107
DB 191 ISKHIDNQLSTSENKIQQLLVHLLIILKHSQPEEDWSTDTSITIAKCKIDINETLG 250
QY 108 YRLSSRSFLWKSYPFRFLDY-LEAFGL-----LSDFLDH-----QAVI 146

; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 615 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-494-106-6

Query Match 4.8%; Score 90; DB 4; Length 654;
Best Local Similarity 20.1%; Pred. No. 0.53;
Matches 83; Conservative 61; Mismatches 152; Indels 116; Gaps 20;

QY 5 ERTLDPKYILKIAL-----KLRQSLSLFFQNSQSLQRAYST-PYSYRIILQKENKEQA 59
DB 129 ERLVHPNQLDILIEDENLNNEASVOFRNDLNNSAAMTFALSFOHYTMRNERAPLYNL 188
QY 60 ARHKCISILEFFKNLLFVHLLSKNOREGCSQTDMAVSTPFPFNRLWYLLSSRSFLWK 119
DB 189 IKNANDSVLRSEQSVIEGHPGLHGAKLKGMNAENFWYSSEFGNVILHRAVFIHKSISR 248
QY 120 SYCPFRFDLYLEAFGLSLDPLDHOAVIK-----FFELETHF-----SYYPVSGFVAP 166
DB 249 -----IOSSNVDVNNVAVKQMPFDLIKLEKEFGEDFNQLQNYHLM--MVHP 291
QY 167 HGYLSLLQDRY-----FFIASVMRTLDKDNFSLTPDL-----IHDLAG 204
DB 292 WQIKHLOSQDYDELBENKLLILSNHSPYIAGLSFRTLLVPKQDPSHIKLSNVH-ITG 350
QY 205 HYPWLLHPSFEFFINNGRLFTKVIKQVALPSKKQRIQTLQSNLIAIVRCFWFTVESGL 264
DB 351 EIRTL-----SEQTYNGPLVTLQILREIM---SKDEDFSHYOST-----Y 387
QY 265 IENHEGRKAYG-----AVLISSPQELG-----HAFIDN--VRVLPLELDQIRPFT 310
DB 388 IDENAGIHFNNDNEAIQTDREQLGTLFRNNLYQFISNETVPVIFSSL--VATVPYNT 445
```

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.ra1

Page 12

QY 311 STPOETL-----FSIRHDELVELTSKLEMMIDQ-----GILLESPIYNO 350  
Db 446 EAPICLTITTYONTYQYKNYEBAK-----OMITDYSKALGLV--IPLYSK 490

RESULT 19  
US-08-911-853-35  
Sequence 35, Application US/08911853  
Patent No. 6048710  
GENERAL INFORMATION:  
APPLICANT: Geritsee, Gijbert  
APPLICANT: Quax, Wilhelmus J.  
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED  
TITLE OF INVENTION: EXPRESSION LEVELS  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genencor International  
STREET: 925 Page Mill Road  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1013  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/911.853  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/699, 092  
FILING DATE: 16-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Glaister, Debra J  
REGISTRATION NUMBER: 33,888  
REFERENCE/DOCKET NUMBER: GC361-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-845-6504  
TELEFAX: 650-846-7620  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 392 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-911-853-35

Query Match 4.7%; Score 89; DB 3; Length 392;

Best Local Similarity 22.4%; Pred. No. 0.31;

Matches 71; Conservative 40; Mismatches 104; Indels 102; Gaps 16;

QY 90 CST-DMAVSTPPFNENLWYRLSSRFSLMKSYCPFPDYLEAFGLSDPFDHOAVIKF 148  
Db 98 CSAGDHVIVRSYVF-----GSTISLFEKYLKRFGEIV--DYPLADLAWQAFAFK- 145  
QY 149 FELETHPSYVSGFVAPHQYLSLDQRYFPFIAVM-----RTLDKNFSLT 195  
Db 146 -----PNTKLLFVESPSPNPLAEIVDIGALAEIAHARGALLAVDNCFCF 188  
QY 196 PDLIHDLGHVPLHLHPSFSEFFINMGRLEFTYVI-----EKV----- 232  
Db 189 PALQOPALAGADVMH--SATKFIGGRLGGVAVAGRAQMEQVGFRTAGPTLSPFNA 247  
QY 233 -----QALPSKKORIOTLOSNI LAIVRCFWFVESGLE-----NHEGRK-----AY 274  
Db 248 WMFLKGLFETLRIMQASASALELAK--WLETPGIDRYVYAGLPSHPQHELAKQOQSAF 305  
QY 275 GAVL-----ISSPQELGHAFIDNVRVPL-----ELDQIIRLPFNTS-----TPQETLS-IR 321  
Db 306 GAVLSFEVKGKGAAMRFIDATRVISITTNIGDTYTTIAHPATTSHGRLSPOBRASAGIR 365

QY 322 HFDELVELTSKLEMMID 338  
Db 366 --DNLVAVAGLEDVVD 380

RESULT 20  
US-09-479-409-35  
Sequence 35, Application US/09479409  
Patent No. 6225106  
GENERAL INFORMATION:  
APPLICANT: Geritsee, Gijbert  
APPLICANT: Quax, Wilhelmus J.  
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED  
TITLE OF INVENTION: EXPRESSION LEVELS  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genencor International  
STREET: 925 Page Mill Road  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1013  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/479,409  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/911.853  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Glaister, Debra J  
REGISTRATION NUMBER: 33,888  
REFERENCE/DOCKET NUMBER: GC361-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-845-6504  
TELEFAX: 650-846-7620  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 392 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-479-409-35

Query Match 4.7%; Score 89; DB 4; Length 392;

Best Local Similarity 22.4%; Pred. No. 0.31;

Matches 71; Conservative 40; Mismatches 104; Indels 102; Gaps 16;

QY 90 CST-DMAVSTPPFNENLWYRLSSRFSLMKSYCPFPDYLEAFGLSDPFDHOAVIKF 148  
Db 98 CSAGDHVIVRSYVF-----GSTISLFEKYLKRFGEIV--DYPLADLAWQAFAFK- 145  
QY 149 FELETHPSYVSGFVAPHQYLSLDQRYFPFIAVM-----RTLDKNFSLT 195  
Db 146 -----PNTKLLFVESPSPNPLAEIVDIGALAEIAHARGALLAVDNCFCF 188  
QY 196 PDLIHDLGHVPLHLHPSFSEFFINMGRLEFTYVI-----EKV----- 232  
Db 189 PALQOPALAGADVMH--SATKFIGGRLGGVAVAGRAQMEQVGFRTAGPTLSPFNA 247  
QY 233 -----QALPSKKORIOTLOSNI LAIVRCFWFVESGLE-----NHEGRK-----AY 274  
Db 248 WMFLKGLFETLRIMQASASALELAK--WLETPGIDRYVYAGLPSHPQHELAKQOQSAF 305  
QY 275 GAVL-----ISSPQELGHAFIDNVRVPL-----ELDQIIRLPFNTS-----TPQETLS-IR 321  
Db 306 GAVLSFEVKGKGAAMRFIDATRVISITTNIGDTYTTIAHPATTSHGRLSPOBRASAGIR 365  
QY 322 HFDELVELTSKLEMMID 338

```
Db 366 --DNLVRVAVGLEDDVD 380
|||: |||: |||:
RESULT 21
US-09-479-453-35
; Sequence 35, Application US/09479453
; Patent No. 6313283
; GENERAL INFORMATION:
; APPLICANT: Gerritse, Gijbert
; APPLICANT: Quax, Wilhelms J.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; TITLE OF INVENTION: EXPRESSION LEVELS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/479,453
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/911,853
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; REFERENCE/DOCKET NUMBER: GC361-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-479-453-35
Query Match 4.7%; Score 89; DB 4; Length 392;
Best Local Similarity 22.4%; Pred. No. 0.31;
Matches 71; Conservative 40; Mismatches 104; Indels 102; Gaps 16;
Qy 90 CST-DNAVSTPFFNNRWLLSSRFLSKSCYCPFFLDYLEAFGLLSDPLDHOAVIKF 148
Db 98 CSAGDHVLSRSVF-----GSTLSLFKYLKRFGLV--DTPPLADLDADAWQAFAK- 145
Qy 149 FELETHFSYYPVSGFVAPHQVLSLQDRYPPIASVM-----RTLDKDNFSLT 195
Db 146 -----ENTKLLFVSPSPNPLAELVDIGALAEIAHARGALLAVDNCFT 188
Qy 196 PDLIHLLGHVPLWLLHPSFSFFFINMGRLFKVI-----EKV----- 232
Db 189 PALQQPALGADVMWH-SATKFDGQGRGLGGVVGARRAQMQVVGFLRTAGPTLSPFNA 247
Qy 233 ----QALPSKKORIOTLOSNIAlAIVRCFWFTVESGLIE-----NHGRK----AY 274
Db 248 WMFLKGLTLRLRMQAQASASALELAR--WLETQGDIDRVVYAGLPSHPOHELAKRQOSAF 305
Qy 275 GAVL----ISSPOLGHAFIDNVRVLP-----ELDQIIRLPENTS---TQETLFS-IR 321
Db 306 GAVLSFVKGKGAARWIFATRVISITNGLDKTTIAHPATTSHGRLSQERASAGIR 365
Qy 322 HFDELVELTSKLEWMLD 338
|||: |||: |||:
Db 366 --DNLVRVAVGLEDDVD 380
|||: |||: |||:
RESULT 22
US-08-996-338-22
; Sequence 22, Application US/08996338
; Patent No. 6087116
; GENERAL INFORMATION:
; APPLICANT: TORIGOE, Kakuii
; APPLICANT: OKURA, Takanori
; APPLICANT: KURIMOTO, Masashi
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/996,338
; FILING DATE: 22-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 74,697/1997
; FILING DATE: 12-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 215,488/1997
; FILING DATE: 28-JUL-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 291,837/1997
; FILING DATE: 09-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TORIGOE-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 309
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-996-338-22
Query Match 4.6%; Score 87.5; DB 3; Length 309;
Best Local Similarity 21.3%; Pred. No. 0.32;
Matches 68; Conservative 50; Mismatches 106; Indels 95; Gaps 19;
Qy 9 DPKYI-----LXIALKLPQSLSLFONSQSLORAVSTPYVYRIILQKENKEQALAHKC 64
Db 15 EFPYKHCSCSLAHEIETTTKSWYKSGSQEHVELNPRSSRIAL-----HDC 62
Qy 65 ISILEFP-----KNLLFVHLLSLKNOREGCSTDMAVVS-----TPFEN-- 103
Db 63 --VLSEFVPELVNDTGSYFFQMKNYTKWKLNIWRNKHSCTERQVTSKIVEVKKPFQIT 120
Qy 104 -ENLYRLLSRFLSKWSCYCPFFLDYLEAFGLLSDPLDHOAVIKFLELTHFSYYPVSG 162
Db 121 CENSYQTLVNSTSLYKN-CKLLLENNK-----NPTIKNAEFDDQ-GYYSVCH 168
Qy 163 FVAPHQVLSLQDRYPPIASVMRTLDKDNFSLTPLDIHLLGHVPLWLLHPSFSFEFFNMG 222
Db 169 FL--HH-----NGKLFNITK-----TFNIT--IVEDRSNIVPVLGPKLNHVAVELG 211
Qy 223 RLFTKVIKQVQALPSKKORIOTLOSNIAlAIVRCFW--FTVESGLIEN-HEGRKAYGAVLS 280
```

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.ra1

Page 14

Db 212 K--NVRNCSALNDEVDY-----YMFGEENGSDPNIHEKE---MIM 251  
QY 281 SPOELGHAFIDNVRVLPLE 299  
Db 252 TPEGKWAH---SKVLRIE 266

RESULT 23  
US-08-996-338-20  
Sequence 20, Application US/08996338  
Patent No. 6087116  
GENERAL INFORMATION:  
APPLICANT: TORIGOE, Kakuji  
APPLICANT: OKURA, Takao  
APPLICANT: KURIMOTO, Mutsaishi  
TITLE OF INVENTION: POLYPEPTIDES  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/996,338  
FILING DATE: 22-DEC-1997  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 74,697/1997  
FILING DATE: 12-MAR-1997  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 215,488/1997  
FILING DATE: 28-JUL-1997  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 291,837/1997  
FILING DATE: 09-OCT-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: TORIGOE-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 521  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-996-338-20

Query Match 4.6%; Score 87.5; DB 3; Length 521;  
Best Local Similarity 21.3%; Pred. No. 0.7; Indels 95; Gaps 19;  
Matches 68; Conservative 50; Mismatches 106;

QY 9 DPKYI-----LKIALKRLSLSLFONOSLOPRAVSTPYRYRIILQENKEKQALARKC 64  
Db 15 EPPYLKHCSCSLAHEIETTTKSMYKSSGQEHVELNPRSSRIAL-----HDC 62  
QY 65 ISILEFP-----KNLLFVHLISLKNQREGCSTMAVVS-----TPPFN-- 103  
Db 63 --VLEFPVPELNDGTSYFQOMKNYQKMKLVIRNKGSCPTEROVTSKIVEYKGFQIT 120  
QY 104 -RLMWRLLSSRSLMKSYCPRPFDYLEAFGLSDPLDHOAVIKFPELETHPSYYPVSG 162  
Db 121 CENSYYQTLVNSTSLYKN-CKKLLLENNK-----NPTIKQAEFEDQ-GYYSQVH 168

QY 163 FVAPHOYLSLQDRYEPFASVWRTLDKONFSLTPDLIHDLGHVPHLHPSSEPFIMNG 222  
Db 169 PL--HH-----NGKLFNITK-----TFNIT--IYEDKSNIVPVILGPKMHNVAVEIG 211  
QY 223 RLFTKVIKQVQALPSKKQRIQTLQSNLIAVRCW--FTVESGLIEN-HEGKRAYGAVLIS 280  
Db 212 K--NVRNCSALNDEVDY-----YMFGEENGSDPNIHEKE---MIM 251  
QY 281 SPOELGHAFIDNVRVLPLE 299  
Db 252 TPEGKWAH---SKVLRIE 266

RESULT 24  
US-08-604-333-2  
Sequence 2, Application US/08604333  
Patent No. 576731  
GENERAL INFORMATION:  
APPLICANT: Parner, Patricia et al.  
TITLE OF INVENTION: Receptor Designated 2P1  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kathryn A. Anderson, Immunex Corporation  
STREET: 51 University Street  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple 7.1  
SOFTWARE: Microsoft Word, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/604,333  
FILING DATE: 21-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Anderson, Kathryn A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2619  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 587-0430  
TELEFAX: (206) 233-0644  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 541 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-604-333-2

Query Match 4.6%; Score 87.5; DB 1; Length 541;  
Best Local Similarity 21.3%; Pred. No. 0.75; Indels 95; Gaps 19;  
Matches 68; Conservative 50; Mismatches 106;

QY 9 DPKYI-----LKIALKRLSLSLFONOSLOPRAVSTPYRYRIILQENKEKQALARKC 64  
Db 34 EPPYLKHCSCSLAHEIETTTKSMYKSSGQEHVELNPRSSRIAL-----HDC 81  
QY 65 ISILEFP-----KNLLFVHLISLKNQREGCSTMAVVS-----TPPFN-- 103  
Db 82 --VLEFPVPELNDGTSYFQOMKNYQKMKLVIRNKGSCPTEROVTSKIVEYKGFQIT 139  
QY 104 -RLMWRLLSSRSLMKSYCPRPFDYLEAFGLSDPLDHOAVIKFPELETHPSYYPVSG 162  
Db 140 CENSYYQTLVNSTSLYKN-CKKLLLENNK-----NPTIKQAEFEDQ-GYYSQVH 187  
QY 163 FVAPHOYLSLQDRYEPFASVWRTLDKONFSLTPDLIHDLGHVPHLHPSSEPFIMNG 222  
Db 168 FL--HH-----NGKLFNITK-----TFNIT--IYEDRSNIVPVILGPKMHNVAVEIG 230

QY 223 RLFTKVIKQVALPSKKQRIQTQLQSNLIAIVRCFW-FTVESGLIEN-HEGRKAYGAVLIS 280  
DB 231 K---NVRLNCALSALNEDVFI-----YMFGEENGSDPNTHIEEKE-----MRIM 270  
QY 281 SPQELGHAFIDNVRVLPLE 299  
DB 271 TPEGKWHIA-----SKVLRIE 285

RESULT 25  
US-09-110-618-2  
; Sequence 2, Application US/09110618  
; Patent No. 6090918  
; GENERAL INFORMATION:  
; APPLICANT: Parinet, Patricia et al.  
; TITLE OF INVENTION: Receptor Designated 2F1  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kathryn A. Anderson, Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Apple 7.1  
; SOFTWARE: Microsoft Word, Version 5.1a  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/110,618  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/604,333  
; FILING DATE: 21-FEB-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Anderson, Kathryn A.  
; REGISTRATION NUMBER: 32,172  
; REFERENCE/DOCKET NUMBER: 2619  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 587-0430  
; TELEFAX: (206) 233-0644  
; TELEX: 756822  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 541 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-110-618-2  
Query Match 4.6%; Score 87.5; DB 3; Length 541;  
Best Local Similarity 21.3%; Pred. No. 0.75;  
Matches 68; Conservative 50; Mismatches 106; Indels 95; Gaps 19;  
QY 9 DPXYI---LKLALQKLSLFPONSQSORAYSTPYVYRIILOKENEKQALARKHC 64  
DB 34 EPFYLKHCSCSLAHEITTKYKWSGSGQEHVELNPRSSRIAL-----HDC 81  
QY 65 ISILEFF-----KNLLFVHLLSLKNQRCGCTDMAVS-----TPFFN-- 103  
DB 82 --VLEFVPELVNDGTGYFFOMKNTQKWLAVIRNRHSCFTEQVTSKIVEVKFFQIT 139  
QY 104 -RNWYRLLSRFLWSKSCYCPREFLDYLEAFGLLSDFLDHQAQVKKFELTHESYYPVSG 162  
DB 140 CENSYYQVLNVTNSLYKN-CKKLLNNK-----NFTIKKNAEFEDQ-GYYSCVH 187  
QY 163 FVAPHQYLSLQDRYFPFIASVMRTDKONFSITPOLIHDLLGHVPLWLLHPSFSEFFINMG 222  
DB 188 FL--HH-----NGKLFNITK-----TFNIT--IVEDRSNIVFVLLGPKLNHVAVELG 230  
QY 223 RLFTKVIKQVALPSKKQRIQTQLQSNLIAIVRCFW-FTVESGLIEN-HEGRKAYGAVLIS 280

DB 231 K---NVRLNCALSALNEDVFI-----YMFGEENGSDPNTHIEEKE-----MRIM 270  
QY 281 SPQELGHAFIDNVRVLPLE 299  
DB 271 TPEGKWHIA-----SKVLRIE 285  
RESULT 26  
US-09-173-151A-28  
; Sequence 28, Application US/09173151A  
; Patent No. 6326472  
; GENERAL INFORMATION:  
; APPLICANT: Timans, Jacqueline C.  
; APPLICANT: Debets, Johannes Eduard Maria  
; APPLICANT: Antonius  
; APPLICANT: Sana, Theodore R.  
; APPLICANT: Bazen, J. Fernando  
; APPLICANT: Kastelein, Robert A.  
; TITLE OF INVENTION: Human Receptor Proteins; Related Reagents and Methods  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DNAX Research Institute  
; STREET: 901 California Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/173,151A  
; FILING DATE: 14-OCT-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/065,776  
; FILING DATE: 17-NOV-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/078,008  
; FILING DATE: 12-MAR-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/081,883  
; FILING DATE: 15-APR-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/095,987  
; FILING DATE: 10-AUG-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/078,416  
; FILING DATE: 18-MAR-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/062,066  
; FILING DATE: 15-OCT-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ching, Edwin P.  
; REGISTRATION NUMBER: 34,090  
; REFERENCE/DOCKET NUMBER: DX0767X  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 852-9196  
; TELEFAX: (650) 496-1200  
; INFORMATION FOR SEQ ID NO: 28:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 541 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-173-151A-28  
Query Match 4.6%; Score 87.5; DB 4; Length 541;  
Best Local Similarity 21.3%; Pred. No. 0.75;



SEQUENCE CHARACTERISTICS:  
 LENGTH: 990 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 US-08-392-625-20

Query Match 4.6%; Score 86.5; DB 2; Length 990;  
 Best Local Similarity 23.1%; Pred. No. 2.4;  
 Matches 57; Conservative 38; Mismatches 79; Indels 73; Gaps 13;

QY 3 YCERLDPKYLKIAL-KLRQSLFFQNSQSORAYSTPYSY-----RIIL 49  
 DB 601 FCPRIYKNIILPATWKINSE--FSETENLNRPATIRKWHIPKDVIIAFGDNRLLL 657  
 QY 50 QKENKE-----KQALAHKICISILEFFKKNLFFVHLLSLSKNOREGCGSTDMVYSTPFF-- 102  
 DB 658 NLLNDKHLIILKELKKGIRILSEF-----INESNNER-----MLEIVTPLYKK 703  
 QY 103 -----NRNLWYRLLSRPSLW---KSYCPRFDDYLEAFGL---LSDFLDHO 143  
 DB 704 TSLKEQSFIIIPKRNKHNKDFSIHLSIPKTYQDNFIQDYLLPFIITELKVNFINKF 763  
 QY 144 AVIKFPELETHFSYYPVSGFVAPHQVLSLQ--DRYFFIASVMRTLDKNFSLTPDLI-H 200  
 DB 764 FYIKFKEDEDFIK-----LRLREDEDSQIYSFIKNW-KDYCLNSELVDY 809  
 QY 201 DLLGHVP 207  
 DB 810 SIVDYVP 816

## RESULT 29

US-08-466-961A-20  
 Sequence 20, Application US/08466961A  
 Patent No. 5843709

GENERAL INFORMATION:  
 APPLICANT: Entian, Karl-Dieter  
 APPLICANT: Gtz, Friedrich  
 APPLICANT: Schnell No. 5843709bert  
 APPLICANT: Augustin, Johannes  
 APPLICANT: Engelke, German  
 APPLICANT: Rosenstein, Ralf  
 APPLICANT: Kaletta, Cortina  
 APPLICANT: Klein, Cora  
 APPLICANT: Wieland, Bernd  
 APPLICANT: Kupke, Thomas  
 APPLICANT: Kellner, Roland  
 TITLE OF INVENTION: Biosynthetic Process for the Preparation of  
 TITLE OF INVENTION: Chemical Compounds  
 NUMBER OF SEQUENCES: 42  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Sterne, Kessler, Goldstein & Fox  
 STREET: 1100 New York Avenue, NW  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: U.S.A.  
 ZIP: 20005  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/466,961A  
 FILING DATE: 06-JUN-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/392,625  
 FILING DATE: 22-FEB-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/876,791  
 FILING DATE: 30-APR-1992

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/784,234  
 FILING DATE: 31-OCT-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Esmond, Robert W.  
 REGISTRATION NUMBER: 32,893  
 REFERENCE/DOCKET NUMBER: 0652.0980004  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 371-2600  
 TELEFAX: (202) 371-2540  
 INFORMATION FOR SEQ ID NO: 20:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 990 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 US-08-466-961A-20  
 Query Match 4.6%; Score 86.5; DB 2; Length 990;  
 Best Local Similarity 23.1%; Pred. No. 2.4;  
 Matches 57; Conservative 38; Mismatches 79; Indels 73; Gaps 13;  
 QY 3 YCERLDPKYLKIAL-KLRQSLFFQNSQSORAYSTPYSY-----RIIL 49  
 DB 601 FCPRIYKNIILPATWKINSE--FSETENLNRPATIRKWHIPKDVIIAFGDNRLLL 657  
 QY 50 QKENKE-----KQALAHKICISILEFFKKNLFFVHLLSLSKNOREGCGSTDMVYSTPFF-- 102  
 DB 658 NLLNDKHLIILKELKKGIRILSEF-----INESNNER-----MLEIVTPLYKK 703  
 QY 103 -----NRNLWYRLLSRPSLW---KSYCPRFDDYLEAFGL---LSDFLDHO 143  
 DB 704 TSLKEQSFIIIPKRNKHNKDFSIHLSIPKTYQDNFIQDYLLPFIITELKVNFINKF 763  
 QY 144 AVIKFPELETHFSYYPVSGFVAPHQVLSLQ--DRYFFIASVMRTLDKNFSLTPDLI-H 200  
 DB 764 FYIKFKEDEDFIK-----LRLREDEDSQIYSFIKNW-KDYCLNSELVDY 809  
 QY 201 DLLGHVP 207  
 DB 810 SIVDYVP 816  
 RESULT 30  
 US-08-645-193B-15  
 Sequence 15, Application US/08645193B  
 Patent No. 5962253  
 GENERAL INFORMATION:  
 APPLICANT: Kupke, Thomas  
 APPLICANT: Gotz, Friedrich  
 APPLICANT: Kempter, Christoph  
 APPLICANT: Jung, Gunther  
 TITLE OF INVENTION: Oxidative Decarboxylation of Peptides  
 TITLE OF INVENTION: Catalyzed by Flavoprotein EpiD  
 NUMBER OF SEQUENCES: 70  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
 STREET: 1100 New York Avenue, Suite 600  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: U.S.A.  
 ZIP: 20005  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/645,193B  
 FILING DATE: 13-MAY-1996  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Esmond, Robert W.  
 REGISTRATION NUMBER: 32,893

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.ra1

Page 18

```
REFERENCE/DOCKET NUMBER: 0652.1540000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 990 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: not relevant
MOLECULE TYPE: protein
us-08-645-193b-15
```

Query Match 4.6%; Score 86.5; DB 2; Length 990;

Best local similarity 21.7%; Pred. No. 2.4;

Matches 62; Conservative 43; Mismatches 100; Indels 81; Gaps 14;

```
QY 3 YCERLDPKYLKIAL-KLRQSLSPFONSQLOAVST-----PYSYRII 48
DB 601 FCPRIIYKNIILKPATWIKNSM--FSETEMLNRPATIREKWHIPKDVIIAFQDNRL 657
QY 49 LOKENE-----KQALAHKCSILEPEKNIILFVHLISLKNRECSIDMAVSTPPF 102
DB 658 LNLNDKHIILIKELKHGRILLESF-----INESNER-----MEIVTPIYK 703
QY 103 -----NNLMYRLSSRFSLM--KSYCPREFLDYLEAFGL--LSDFLDH 142
DB 704 KTSLEQSFIIKNNKKNFNNLKDWPFSIHLSIPTQDNFTQDYLPTTELKANNFLNK 763
QY 143 QAVIKFELETHFSYVSGVAPHPQVLSLIQ--DRYPIASVMTLKDKNFSLTPLDI 199
DB 764 FVYIKKEDEDFIK-----LRLREDEDSQSYSTIKW--DQYCLNLSLYD 809
QY 200 HDLGHVPLMLH-----PFSSEFIMNGRLFTKVIKQVALPSK 238
DB 810 YSIVDYVEVRYGGPHVIEDIENFMWDSLSINIIQSEFKIPKE 855
```

RESULT 31

US-09-134-001C-4322

Sequence 4322, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

TITLE OF INVENTION: EPIDERMIDS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: CTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C

PRIOR FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 4322

LENGTH: 656

TYPE: PRP

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4322

Query Match 4.6%; Score 86; DB 4; Length 656;

Best local similarity 22.5%; Pred. No. 1.5;

Matches 51; Conservative 26; Mismatches 78; Indels 70; Gaps 11;

```
QY 4 CERTLDPK---YILKIALKQSLSPFONSQLOAVS-----TPYS-----YRI 47
DB 369 CFTVPNNPAPALAEKGVNDKLLSFQSEKLRMSFLMKGLYLPINGNLIHGC 428
QY 48 LOKENEQA-----LAHAKCSILEPEKNIILFVHLISLKNRECSIDMAVSTP 100
DB 429 IPVDENGEMSEFIEGERLSGELLDVEFYVRRAPFH-----KESTEDISTDL----- 477
QY 101 FPNRLMYRLSSRFSLMKSYCPREFLDYLEAFGLSLDFLDHQAIVIKFELETHFSYVP 160
```

```
DB 478 -----VWY-----LMT-----GKYSLSGKQAMTF---EYFIEDKA 507
QY 161 SGFAHPHQVLSLIQDRYPIASVMTLKDKNFSLTPLDIHLGHVP 207
DB 508 SHKEKNPYTYLRKD-----VDMIRKMLKD-FGLNPDEGRITNGHTP 548
```

RESULT 32

US-09-810-347-4

Sequence 4, Application US/09810347

Patent No. 6461847

GENERAL INFORMATION:

APPLICANT: YE, Jane et al.

TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC

TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES

TITLE OF INVENTION: THEROP

FILE REFERENCE: CL001169

CURRENT APPLICATION NUMBER: US/09/810,347

CURRENT FILING DATE: 2001-03-19

NUMBER OF SEQ ID NOS: 6

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 4

LENGTH: 360

TYPE: PRP

ORGANISM: Human

US-09-810-347-4

Query Match 4.5%; Score 84.5; DB 4; Length 360;

Best local similarity 19.6%; Pred. No. 0.85;

Matches 67; Conservative 51; Mismatches 117; Indels 107; Gaps 14;

```
QY 45 YRIILOKENEKAQALAHKCSILEPEKNIILFVHLISLKNR---EGCSTDMAVSTP 100
DB 30 FRGRIQOERKN---IRPNIIIVLTDQDVELGSLQVANNKTRKMEHGATPFINAVTTP 85
QY 101 F-----FNNLMYRLSSRFSLMKSYCPREFLDYLEAFGLSLDFLDHQAIV 145
DB 86 MCCPSRSSMLTGKYVHNHVVYNNNENCSPSQMOMHEHRRPAAVYLVNNGYRTAFG---- 141
QY 146 IKFPELETHFSYVSGVAPHPQVLSLIQDRYPIASVMT--LDKNFSLTGPLDIHDL 203
DB 142 KYLN-EYNSYIP-PGW---REWLGLIKNSRFYNYTCNGGKKEKGFYADYFTDLI 195
QY 204 -----GHVPMILHSFSEFFIMKR----- 223
DB 196 TNESINFPKSKMYHRVVMVAVISHAAPGDESAFOFSKLYPNASQITTSYNAFPM 255
QY 224 -----LTKVIEKVALPSKKORIQTOSNLIAIVRCWFTVESGLIE 266
DB 256 DKHWIMQYTPMLPIHMEFTNIIQ-----RRKLTQMSVDSVRLNMLVETGELE 307
QY 267 NHEGRKAYGAVLSSPOELGAHFIDNVAVLPLELDQIIRLPF 308
DB 308 N-----TYIITYADHGHIQPGGLVKGKSMPEYDD--IRVPF 342
```

RESULT 33

US-09-810-347-2

Sequence 2, Application US/09810347

Patent No. 6461847

GENERAL INFORMATION:

APPLICANT: YE, Jane et al.

TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC

TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES

TITLE OF INVENTION: THEROP

FILE REFERENCE: CL001169

CURRENT APPLICATION NUMBER: US/09/810,347

CURRENT FILING DATE: 2001-03-19

NUMBER OF SEQ ID NOS: 6

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 2

LENGTH: 372



```
; TYPE: PRT
; ORGANISM: Human
; US-09-810-347-2

Query Match      4.5%; Score 84.5; DB 4; Length 372;
Best Local Similarity 19.6%; Pred. No. 0.89;
Matches 67; Conservative 51; Mismatches 117; Indels 107; Gaps 14;

QY 45 YRIILQENKEKQALAHKICISILFFKNNLFLVHLLSKNQR-----ECSTDMAVSTP 100
DB 30 FRGRIQOERKN-----IRPNIILVLDQDVELGSLQWNNKTRKIMEHGGAATINAFVTP 85
QY 101 F-----FNNRLWYLLSSRFSLWKS-YCPRFFLDYLEAFGLSLDFLDHOAV 145
DB 86 MCCPSRSMLTGKYVHNHNVTNNENSCSPQAMHEPRTFAVYLANNTGYRTAFTG----- 141
QY 146 IKFPELETHFSYYPVSGFVAPHQYLSLLQDRYFFIASVMRT--LDKDNFSLPDLIHDLL 203
DB 142 -KYLN-EVNGSYIP-PCW---REWLGLIKNSRFNYTVCRNGIKKHEKGFYADYFTDLI 195
QY 204 -----CHVPWLLHPSFGEFFINMGR----- 223
DB 196 TNESINYFKSKMYPRPVMVISHAAPHGPDSSAPQSKLYPNASQHLTPSYNAPNM 255
QY 224 -----LTKVIEKVOALPSKQRIOTLOSNIATVRCFWFTVBSGLIE 266
DB 256 DKHIMQVTPMLPIHMEFNLIQ-----RKLQTLMSVDDSVRLYNLVEITGELE 307
QY 267 NHEGRKAYGAVLSSPOELGHAFIDNVRVLPLELDQIRLFP 308
DB 308 N-----TYIIVTDHGVHIGQGLVKGKSNPYDFD--IRVFF 342

RESULT 34
5194600-4
; Patent No. 5194600
; APPLICANT: BUSSEY, HOWARD; BOONE, CHARLES; SOMMER, STEVE S.;
; HILL, KATHRYN; MEADEN, PHILIP
; TITLE OF INVENTION: GENES WHICH PARTICIPATE IN B-GLUCAN
; ASSEMBLY AND USE THEREOF
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,316
; FILING DATE: 05-NAR-1990
; SEQ ID NO: 4:
; LENGTH: 1365
5194600-4

Query Match      4.4%; Score 83.5; DB 6; Length 1365;
Best Local Similarity 19.0%; Pred. No. 8.3;
Matches 74; Conservative 49; Mismatches 100; Indels 167; Gaps 19;

QY 89 GCSTDMVSTP--FNNRLWYLLSS--RFLSKVYC----- 122
DB 172 GCPT---VIDSDFEFNKLFHEAMNKGKFRP-IMRSTCLDGKSVYPLTHPLITLQ 227
QY 123 -----PRFLD-----YLEAFGLSLDFLDH--- 142
DB 228 NGRSMSSIPQLKKILYTPKREILVGADNKKQLHLEPEELRELKRLVTSLSIEFYQKXD 287
QY 143 -----QAVIKFF-----ELETHFSYYPVSGF----- 163
DB 288 ITATLNTKSIIVNNPLISKLIKIVSVNNKDIITSNEELNSKGFYNNLMGLIYINGQNWKI 347
QY 164 --VAPHOYLSLLQDRYFFIASVMRTDK-----DNFSLPDLIHDLLGHVPWL----- 209
DB 348 TSLTPNLLTALKTEYQSLKITNLLQLELPSKCIILDSKFLNKNFSQSLKQLONLOPIK 407
QY 210 --LH--PSSEFFINMGRL-----FTKVIKVOAL--PSKQRIOTLOSNIATVRCFWF 258
DB 408 MDLHTIPGSESVIYFNIESDPQYDELVNSVQAFDQKSKFGELPEIKQNWSEII-----F 463
QY 259 TVESGLIENHEGRKAYG-----AVLISS-----POELGHAFIDNVRVLPLELDQIRLPTS 311

; TYPE: PRT
; ORGANISM: Human
; US-09-810-347-2

Query Match      4.4%; Score 83; DB 2; Length 491;
Best Local Similarity 18.4%; Pred. No. 2;
Matches 60; Conservative 48; Mismatches 110; Indels 108; Gaps 12;

QY 15 KIALKROSLSLFPQNSLOSIRAYSTPYPIYRILLOKENKEKQALAHKICISILEFFKNL 74
DB 56 ELAERLQREEREPASSQSQGQSLTFSKFE---EKTNEK---TRKTTVTKKFF--- 105
QY 75 LEVHLLSLSKNQRSGCSTDM--AVSTPFFNRLWYLLSSRFSLWKSYPREFLDYLEA 132
DB 106 -----SASSRAGSKKEIQEAKAPSPINR-----QTSIEDTRVTKEFIDFLKT 148
QY 133 FGLLSDFLDH---QAVIKFFELETHFSYYPVSGFVAPHQYLSLLQDRYFFIASVMRTDK 189
DB 149 F-----HKTGOEVTK--QTKVFLAMPYKRDLSIEBQSECTQDFYQNVARMOT--- 195
QY 190 DNFSUTPOLIHDLLGHVPWLLHPSFGEFFINMGRFLTIVIEKVOALPSKQRIOTLOSNI 249
```

Fri Jan 10 10:56:11 2003

US-09-438-185a-1047.ra1

Page 20

Db 196 -----RGKVP-----PENYEKIM----- 208  
Qy 250 IAIYRCWFVESGLIENHGRKAGAVL---ISSPOLGHAFIDNVRVLPLELDQIIRL 306  
Db 209 -----DQLEKHMTRLYKVFCEETIDDEKDLAQKRIKALHWVTPOMLCV 255  
Qy 307 PFNTSTPOETLFSIRHFDLVELTSG 332  
Db 256 PVNEEIPVSDMVVKAITDIIMDSK 281

RESULT 36  
US-08-913-159-2  
Sequence 2, Application US/08913159  
Patent No. 6300109  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: Plasmid-derived type II  
restriction-modification systems from Lactococcus lactis  
NUMBER OF SEQUENCES: 14  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/913,159  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DK 0179/95  
FILING DATE: 17-FEB-1995  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 284 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-913-159-2

Query Match .4.3%; Score 82; DB 4; Length 284;  
Best Local Similarity 26.1%; Pred. No. 1.1; Mismatches 69; Indels 52; Gaps 12;  
Matches 53; Conservative 29; Mismatches 69; Indels 52; Gaps 12;  
Qy 184 MRLDKNFSITPDL-----IHDLGHVPMILHPSFSF---FINMGRLEPTKYIEKQVL 235  
Db 1 MNLQKXKIKILPFTKMGKQQLPHIOYLMPEKINIFPEPFIGGALPFE-----L 53  
Qy 236 PSKKORIQTLSNLIAIVRCFWFTVES--GLIENH--EGKAYGAVLISSPOLGHAFID 291  
Db 54 APOCAVINDENSELINCYROMKNPQLIELTNNHORENSKEYYLDLRSSDRD--GRID 110  
Qy 292 NVRVLPLELDQIIRLPNTSTPOETLFSIR-HFDLVELTSGLEW-----MLDQ 339  
Db 111 KMS-----EYERARI-----MYMLRVDFNGLYKVNKQFNVPYGRYKNPXIVDK 156  
Qy 340 GLEESIPLY---NOEKYLSG-PE 358  
Db 157 ELIESISEYLNNSIKIMSGDPE 179

RESULT 37  
US-09-022-875-2  
Sequence 2, Application US/09022875  
Patent No. 6071894  
GENERAL INFORMATION:  
APPLICANT: Wallis, Nicola G.  
TITLE OF INVENTION: No. 6071894el Compounds  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert, Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch St  
CITY: Philadelphia

STATE: PA  
COUNTRY: USA  
ZIP: 19103-2793  
COMPUTER READABLE FORM:  
MEDIUM TYPE: diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/022,875  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/039,478  
FILING DATE: 25-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Falk, Stephen T  
REGISTRATION NUMBER: 36,795  
REFERENCE/DOCKET NUMBER: GMS0009  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-994-2488  
TELEFAX: 215-994-2222  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 861 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-022-875-2

Query Match 4.3%; Score 82; DB 3; Length 861;  
Best Local Similarity 22.4%; Pred. No. 6; Mismatches 144; Indels 92; Gaps 22;  
Matches 87; Conservative 66; Mismatches 144; Indels 92; Gaps 22;  
Qy 11 KYILKIALKIKQSLSLFPQNSQSLQRAVSTP---YSYRI--ILQRE-----NREKA 58  
Db 488 KFOYSITKQLYRTDILFQFNDSIKQTYVENMLINAGYQINLOOSITTYVYNOSKYI 547  
Qy 59 LARHKCISILEFFKNILFVHLISLS---KNQRE--GCSIDMAVNSTPFNNMLWYRLLSR 114  
Db 548 -----KTIPLQNHIDNTQOHEQALSWVINKEROAGATTD-----TPGINKV--LIPIG 595  
Qy 115 FSLWMSYCPREFLDYLAFLGLSDPLDQA--VIKFE---LETHFSYVSGFVAPOHY 169  
Db 596 TSPK-----GILA--IDYSSQVINPYDASILEMNLBSLA-----VEN 634  
Qy 170 LSLDORFPFASVMTKDNF--SLTPDL---IHDLGHVPMILHPSFSFIFIMGR 224  
Db 635 VTLKQTRRESMLQEROLTHSNLKSISHDITPTLTITMGMDILVSHKD----- 685  
Qy 225 FTKVIEKVALP--SKKORIQTLSNLIAIVRCFWFTVE-----SGLIEN---HE 269  
Db 686 -MSIEKQQLVHSPQSOYLTYLVTNITSLTKQSSNVQIKQOPYVLSLVEIMILE 744  
Qy 270 GRKAGAVLISSPOLGHAFIDNVRVLPLELDQIIRLPNTSTPOETLFSIRHD-ELVE 328  
Db 745 RHHLKRIITVSSSVNLOPHIDSKLILQALFNLINAVKHTSTDTKINLSIRYASBOIE 804  
Qy 329 LTSKLEWMLDQGLIESIPLYNOEKYLSGF 357  
Db 805 FA-----VIDEG--PGISLEQOKIREPF 826

RESULT 38  
US-09-810-347-5  
Sequence 5, Application US/09810347  
Patent No. 6461847  
GENERAL INFORMATION:  
APPLICANT: Ye, Jane et al.  
TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC  
ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES  
THEROF

FILE REFERENCE: CLO01159  
 CURRENT APPLICATION NUMBER: US/09/810,347  
 CURRENT FILING DATE: 2001-03-19  
 NUMBER OF SEQ ID NOS: 6  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 5  
 LENGTH: 307  
 TYPE: PRT  
 ORGANISM: Human  
 US-09-810-347-5

Query Match 4.3%; Score 81.5; DB 4; Length 307;  
 Best Local Similarity 19.8%; Pred. No. 1.4;  
 Matches 57; Conservative 39; Mismatches 93; Indels 99; Gaps 12;  
 QY 95 AVVSTPF-----FNRNLWYLLSSRFLWKS-YCPRFELDYLEAFGLLSDF 139  
 DB 27 AFVTTMCCPSRSMULTGKYVHNHNYTTNNCCSPSQAMHEPRTFAVYLNNTGYRTAF 86  
 QY 140 LDHQAVIKFPELETHESYYPVSGFVAPHOYLSLLODRYFPIASVMRT--LDKDNFSLTPD 197  
 DB 87 FG-----KYLN-EYNGSYIP-PCW---REWLGLIKNSRFYNTVCRNGIKEKHGFYAKD 136  
 QY 198 LHHDL-----GHVPMLLHPSFSEFFINMGR----- 223  
 DB 137 YFTDLITNESINYFMSKMYPHRPVMMVISHAAPHGEDSAFOFSKLYPNASQHLTPSY 196  
 QY 224 -----LFTKVIEKVALPSKKKRIOTLQSNLIAIVRCFWFTV 260  
 DB 197 NVAPNMDKHWIMQVTGPMPLFIHMEFTNIIQ-----RKRLQTLMSVDSVERLYNMLV 248  
 QY 261 ESSLIEHGRKAYGAVLSSPDELGHAFIDNVRVLPLELDQIIRLPF 308  
 DB 249 ETGELEN-----TYIITADGHYHIGQGLVKGKSMYDFD--IRVPF 289

RESULT 39  
 US-08-766-014-3  
 Sequence 3, Application US/08766014  
 Patent No. 5744312  
 GENERAL INFORMATION:  
 APPLICANT: Mamone, Joseph A.  
 APPLICANT: Davis, Maria  
 APPLICANT: Sha, Dan  
 TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASE  
 NUMBER OF SEQUENCES: 28  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Lyon & Lyon  
 STREET: 633 West Fifth Street  
 CITY: Los Angeles  
 STATE: California  
 COUNTRY: U.S.A.  
 ZIP: 90071-2066  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
 MEDIUM TYPE: storage  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: IBM P.C. DOS 5.0  
 SOFTWARE: FastSeq Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/766,014  
 FILING DATE: Herewith  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 PRIOR APPLICATION DATA: including application  
 PRIOR APPLICATION DATA: described below:  
 APPLICATION NUMBER: 60/008,688  
 FILING DATE: December 15, 1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Warburg, Richard J.  
 REGISTRATION NUMBER: 32,327  
 REFERENCE/DOCKET NUMBER: 223/104

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (213) 489-1600  
 TELEFAX: (213) 955-0440  
 TELEX: 67-3510  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 608 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-08-766-014-3

Query Match 4.3%; Score 81.5; DB 1; Length 608;  
 Best Local Similarity 18.0%; Pred. No. 4;  
 Matches 67; Conservative 57; Mismatches 120; Indels 129; Gaps 18;  
 QY 65 ISILEFFKNLLFVHLLSKNQREGCTDMAVSTPFFNRLMWYLLSSRFLWKSYP 124  
 DB 1 MKLLEFNKLEFFSLIDNIKKE-----SSEIVD-----NHKVEKWSK--- 38  
 QY 125 FFLDYLEAFGLLSDFLDHOAVIKFPELETHESYYPV-----SGFVAPHQYLSLQ 174  
 DB 39 --VDIKELVTLQD-----NRTAFYPLIVEGEIKKIAFSFGKDTVIDVQ 83  
 QY 175 DRYFPIASVMRTDKDNFSLTPDLIHDLL-----GHVPMLLHPSFSEF 217  
 DB 84 TE-----DLKEIFEKEDFEFTTHEIKDFLVLSYKIECKSKYIDTAVMAYLLNPS 138  
 QY 218 FINMGRFTKVIEKVALPS-----KKRIOTLQSNLIA--IVRCF----- 257  
 DB 139 --LDRVLKYL-KVD-VFSYEGIFGKGRDKKIEIDENILADYICSRVYFLDLKEKL 194  
 QY 258 -----FTVESGLIENHEGRKAYGAVLSSP-DELGHAFIDNVRVLPLELDQI 303  
 DB 195 MNFIEEMDKLLLEIMPLVLEVLKMEVSGFTDKVELKESQKIDDRIGEI---LDKI 251  
 QY 304 IR---LPFTSTPQE-TLFSIRHFDLVELTSKLEWMLDOGLLESIPLYN----- 349  
 DB 252 YKEAGYQFNVNSKQJSEFLFEKLNLFVTKTKTGYSTDSEVLQVLPYNDIVSDIIEYR 311  
 QY 350 -----QEKYLSGF 357  
 DB 312 QUTKLKSTYIDGF 324

RESULT 40  
 US-08-493-092-2  
 Sequence 2, Application US/08493092  
 Patent No. 5728807  
 GENERAL INFORMATION:  
 APPLICANT: Shiloh, Yosef  
 APPLICANT: Tagle, Danilo A.  
 APPLICANT: Collins, Francis S.  
 TITLE OF INVENTION: Ataxia-Telangiectasia Gene  
 NUMBER OF SEQUENCES: 7  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Reising, Echington, Barnard & Perry  
 STREET: P.O. Box 4390  
 CITY: Troy  
 STATE: Michigan  
 COUNTRY: US  
 ZIP: 48099  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/493,092  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.ra1

Page 22

```

NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,955
REFERENCE/DOCKET NUMBER: P-310 (TAU)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 689-3500
TELEFAX: (810) 689-4071
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1708 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-493-092-2

Query Match
Best Local Similarity 4.3%; Score 81.5; DB 1; Length 1708;
Matches 61; Conservative 58; Mismatches 114; Indels 97; Gaps 17;

QY 10 PKYILKIAL-----KLROSLSLFFONSOSLQRAVSTPYRYRIIL---OKENKKEQ 57
DB 34 PSHVIAKATPAVINSCHKTKKSLILELSK-----SPDSYQKILLALICEQALETNN 83
QY 58 ALARHKCISILEFFKMLFVHLLSLSKNQRGCTDMAVYSTPEFNNIMYRLSSRFSL 117
DB 84 VYKGRILKTYHLFVSL-----KOIKSLGGAMA-----FVLADVIYTLIH----- 126
QY 118 WKSVCPRFLDY-LEAFGLSDPLDH--QAVIKFE--LETHSYYPVSGFVAP----- 166
DB 127 YINORPSCIMDVSLRSFSLCCDLSCQVCTAVYCKDALENHL--HVIAGTLLPLVYEQV 184
QY 167 ---HOYLSLQDRYFPPIASVMTLKDKNFSLTPDLIHDLGHV----- 206
DB 185 EVQKQVLDL--KYL---VIDNKDNEMLYITIKLDPFPHVFKDLRTQOKIKYSRG 238
QY 207 PMLHPSEFPIINMGRLEFTYIEKVALPSKKORIQTLOSNIATVRCFWFTVESGLIE 266
DB 239 PFSLEIEINH-FLSVSVYDALPLTRLEGLKDLRQLRLKHQWDIMRASQDNPDGIMV 297
QY 267 -----NHEGRK---AYGAVL 278
DB 298 KLVVNLQLSKMAINHTEKEVLEAVGSL 327

RESULT 41
US-08-508-836A-2
; Sequence 2, Application US/08508836A
; Patent No. 5777093
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; APPLICANT: Tagle, Danilo A.
; APPLICANT: Collins, Francis S.
; TITLE OF INVENTION: Ataxia-telangiectasia Gene
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reising, Ethington, Barnard & Perry
; STREET: P.O. Box 4390
; CITY: Troy
; STATE: Michigan
; COUNTRY: US
; ZIP: 48099
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/508,836A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: P-313 (TAU)
```

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 689-3500
TELEFAX: (810) 689-4071
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1708 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-508-836A-2

Query Match
Best Local Similarity 4.3%; Score 81.5; DB 1; Length 1708;
Matches 61; Conservative 58; Mismatches 114; Indels 97; Gaps 17;

QY 10 PKYILKIAL-----KLROSLSLFFONSOSLQRAVSTPYRYRIIL---OKENKKEQ 57
DB 34 PSHVIAKATPAVINSCHKTKKSLILELSK-----SPDSYQKILLALICEQALETNN 83
QY 58 ALARHKCISILEFFKMLFVHLLSLSKNQRGCTDMAVYSTPEFNNIMYRLSSRFSL 117
DB 84 VYKGRILKTYHLFVSL-----KOIKSLGGAMA-----FVLADVIYTLIH----- 126
QY 118 WKSVCPRFLDY-LEAFGLSDPLDH--QAVIKFE--LETHSYYPVSGFVAP----- 166
DB 127 YINORPSCIMDVSLRSFSLCCDLSCQVCTAVYCKDALENHL--HVIAGTLLPLVYEQV 184
QY 167 ---HOYLSLQDRYFPPIASVMTLKDKNFSLTPDLIHDLGHV----- 206
DB 185 EVQKQVLDL--KYL---VIDNKDNEMLYITIKLDPFPHVFKDLRTQOKIKYSRG 238
QY 207 PMLHPSEFPIINMGRLEFTYIEKVALPSKKORIQTLOSNIATVRCFWFTVESGLIE 266
DB 239 PFSLEIEINH-FLSVSVYDALPLTRLEGLKDLRQLRLKHQWDIMRASQDNPDGIMV 297
QY 267 -----NHEGRK---AYGAVL 278
DB 298 KLVVNLQLSKMAINHTEKEVLEAVGSL 327

RESULT 42
US-08-508-836A-8
; Sequence 8, Application US/08508836A
; Patent No. 5777093
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; APPLICANT: Tagle, Danilo A.
; APPLICANT: Collins, Francis S.
; TITLE OF INVENTION: Ataxia-telangiectasia Gene
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reising, Ethington, Barnard & Perry
; STREET: P.O. Box 4390
; CITY: Troy
; STATE: Michigan
; COUNTRY: US
; ZIP: 48099
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/508,836A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,955
; REFERENCE/DOCKET NUMBER: P-313 (TAU)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 689-4071
```

/ INFORMATION FOR SEQ ID NO: 8:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 3056 amino acids  
/ TYPE: amino acid  
/ STRANDEDNESS: single  
/ TOPOLOGY: linear  
US-08-508-836A-8

Query Match 4.3%; Score 81.5; DB 1; Length 3056;  
Best Local Similarity 18.5%; Pred. No. 47;  
Matches 61; Conservative 58; Mismatches 114; Indels 97; Gaps 17;

QY 10 PKYIKIAL-----KLRQSLFFQNSQSLQRAYSTPYSYRIL---QKENKEQ 57  
DB 1382 PSHVIKATPAYISNCHTKLKSLLELSK-----SPDSYQKILLAIACEQAETNN 1431  
QY 58 ALARHKCISILEFFKNLLFVHLLSLSKNOREGCTDMAVSTPFFNRNLYRLLSRFSL 117  
DB 1432 VYKXRIKILYHLFVSLIL-----KDKSLGGAWA-----FVLRDVIYTLIH----- 1474  
QY 118 WKSVCPRFFLDY-LEAFGLLSDFLDH--QAVIKFFE--LETHFSYYPVSGFVAP----- 166  
DB 1475 YINQRPSCIMDVSLRSFSLCCDLLSQVCQTAVTYCKDALENHL--HVIVGTLPVYEQV 1532  
QY 167 ---HQYLSLLQDRYFFPIASVMRTLDKDNFSLTPDLIHDLGHV----- 206  
DB 1533 EVQKQVLDLL--KYL-----VIDNKNENLYITIKLLDPFDPHVFKDLRITQOKIKYSG 1586  
QY 207 PWLLHPSEFFINMGRLEFTKVIKQVQALPSKQRIOTLQSNLIAIVRCFWFTVESGLIE 266  
DB 1587 PFSLEEBINH-FLSVSVYDALPLTRLEGLKDLRQLELHKDQVDMINRASQDNPQDGINV 1645  
QY 267 -----NHEGRK-----AYGAVL 278  
DB 1646 KLVNLLQLSKMAINHTEGEVLEAVGSCL 1675

## RESULT 43

US-08-629-001A-3  
/ Sequence 3, Application US/08629001A  
/ Patent No. 5858661  
/ GENERAL INFORMATION:  
/ APPLICANT: Shiloh, Yosef  
/ TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE AND ITS  
/ TITLE OF INVENTION: GENOMIC ORGANIZATION  
/ NUMBER OF SEQUENCES: 139  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: Kohn & Associates  
/ STREET: 30500 No. 5858661thwestern Hwy.  
/ CITY: Farmington Hills  
/ STATE: Michigan  
/ COUNTRY: US  
/ ZIP: 48334  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Floppy disk  
/ COMPUTER: IBM PC compatible  
/ OPERATING SYSTEM: PC-DOS/MS-DOS  
/ SOFTWARE: Patentin Release #1.0, Version #1.30  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/08/629,001A  
/ FILING DATE:  
/ CLASSIFICATION: 435  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Kohn, Kenneth I.  
/ REGISTRATION NUMBER: 30,955  
/ REFERENCE/DOCKET NUMBER: 2290.00032  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: (810) 539-5050  
/ TELEFAX: (810) 539-5055  
/ INFORMATION FOR SEQ ID NO: 3:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 3056 amino acids  
/ TYPE: amino acid

/ STRANDEDNESS: single  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: protein  
/ ORIGINAL SOURCE:  
/ ORGANISM: Homo sapiens  
US-08-629-001A-3

Query Match 4.3%; Score 81.5; DB 2; Length 3056;  
Best Local Similarity 18.5%; Pred. No. 47;  
Matches 61; Conservative 58; Mismatches 114; Indels 97; Gaps 17;

QY 10 PKYIKIAL-----KLRQSLFFQNSQSLQRAYSTPYSYRIL---QKENKEQ 57  
DB 1382 PSHVIKATPAYISNCHTKLKSLLELSK-----SPDSYQKILLAIACEQAETNN 1431  
QY 58 ALARHKCISILEFFKNLLFVHLLSLSKNOREGCTDMAVSTPFFNRNLYRLLSRFSL 117  
DB 1432 VYKXRIKILYHLFVSLIL-----KDKSLGGAWA-----FVLRDVIYTLIH----- 1474  
QY 118 WKSVCPRFFLDY-LEAFGLLSDFLDH--QAVIKFFE--LETHFSYYPVSGFVAP----- 166  
DB 1475 YINQRPSCIMDVSLRSFSLCCDLLSQVCQTAVTYCKDALENHL--HVIVGTLPVYEQV 1532  
QY 167 ---HQYLSLLQDRYFFPIASVMRTLDKDNFSLTPDLIHDLGHV----- 206  
DB 1533 EVQKQVLDLL--KYL-----VIDNKNENLYITIKLLDPFDPHVFKDLRITQOKIKYSG 1586  
QY 207 PWLLHPSEFFINMGRLEFTKVIKQVQALPSKQRIOTLQSNLIAIVRCFWFTVESGLIE 266  
DB 1587 PFSLEEBINH-FLSVSVYDALPLTRLEGLKDLRQLELHKDQVDMINRASQDNPQDGINV 1645  
QY 267 -----NHEGRK-----AYGAVL 278  
DB 1646 KLVNLLQLSKMAINHTEGEVLEAVGSCL 1675

## RESULT 44

US-08-874-266-2  
/ Sequence 2, Application US/08874266  
/ Patent No. 5955279  
/ GENERAL INFORMATION:  
/ APPLICANT: Gatti, Richard A.  
/ TITLE OF INVENTION: ATAXIA-TELANGIECTASIA MUTATIONS IN THE ATM GENE  
/ NUMBER OF SEQUENCES: 33  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: Knobbe, Martens, Olson and Bear  
/ STREET: 620 Newport Center Drive 16th Floor  
/ CITY: Newport Beach  
/ STATE: CA  
/ COUNTRY: USA  
/ ZIP: 92660  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Diskette  
/ COMPUTER: IBM Compatible  
/ OPERATING SYSTEM: DOS  
/ SOFTWARE: FastSeq for Windows Version 2.0  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/08/874,266  
/ FILING DATE:  
/ CLASSIFICATION: 435  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER:  
/ FILING DATE:  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Ways Vensko, Nancy  
/ REGISTRATION NUMBER: 36,298  
/ REFERENCE/DOCKET NUMBER: UCLA006.006A  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: 619-235-8550  
/ TELEFAX: 619-235-0176  
/ TELEX:  
/ INFORMATION FOR SEQ ID NO: 2:  
/ SEQUENCE CHARACTERISTICS:

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.ra1

Page 24

LENGTH: 3056 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-08-674-266-2

Query Match  
Best Local Similarity 4.3%; Score 81.5; DB 2; Length 3056;  
Pred. No. 47;  
Matches 61; Conservative 58; Mismatches 114; Indels 97; Gaps 17;

QY 10 PKYILKIAL-----KLRQSLSLFQNSQSLQRAVSPYVYRIIL--QKENKQ 57  
DB 1382 PSHVIAKATPAVYISNCHKTKLSILEILSK-----SPDSYQKILALICEQALETNN 1431  
QY 58 ALARHKCISILEFFKNLLEFVHLISLKNQREGCTDMAVSTPEFNRNLMYLLSRSFSL 117  
DB 1432 VYKGRILKIYHLFVSL-----KDKSGLGAWA-----FVLADVIYTLIH----- 1474  
QY 118 WKSYCPRFPLDY-LEAFGLSDPLDH--QAVIKFEE--LETHSYVPSGVAP----- 166  
DB 1475 YINQRPSCIMDVLSRFSCLCDLISQVCQTAVYCKKALENHL--HVIYGTLLPLVYEQV 1532  
QY 167 ---HOYLSLLQDRYFPPIASVWRTLDKNFSLTPDLIHDLLGHV----- 206  
DB 1533 EVQKQVLDL--KYL-----VIDKNQENNYITIKLDPFPHVFKDLRTQOKIKYSRG 1586  
QY 207 PMLHPSEFPIINMGRLEFTKVIKQVQALPSKKORIQTLSNLIAYRCFWFVSGGLE 266  
DB 1587 PSLLEBINH-FLSVSYDALPULTRBGLKDLRQLELHKQWDIMRASQDNPDGIMV 1645  
QY 267 -----NHEGRK---AYGAVL 278  
DB 1646 KLVVNLQLSKMAINHTEKEVLEAVGSCL 1675

RESULT 45  
US-08-642-274D-3  
Sequence 3, Application US/08642274D  
Patent No. 6200749  
GENERAL INFORMATION:  
APPLICANT: Shiloh, Yosef  
TITLE OF INVENTION: MUTATED FORMS OF THE ATAXIA-TELANGIECTASIA GENE AND METHOD TO  
FILE REFERENCE: 229000033  
CURRENT APPLICATION NUMBER: US/08/642,274D  
CURRENT FILING DATE: 1996-05-03  
NUMBER OF SEQ ID NOS: 220  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 3  
LENGTH: 3056  
TYPE: PKT  
ORGANISM: Homo sapiens  
US-08-642-274D-3

Query Match  
Best Local Similarity 4.3%; Score 81.5; DB 4; Length 3056;  
Pred. No. 47;  
Matches 61; Conservative 58; Mismatches 114; Indels 97; Gaps 17;

QY 10 PKYILKIAL-----KLRQSLSLFQNSQSLQRAVSPYVYRIIL--QKENKQ 57  
DB 1382 PSHVIAKATPAVYISNCHKTKLSILEILSK-----SPDSYQKILALICEQALETNN 1431  
QY 58 ALARHKCISILEFFKNLLEFVHLISLKNQREGCTDMAVSTPEFNRNLMYLLSRSFSL 117  
DB 1432 VYKGRILKIYHLFVSL-----KDKSGLGAWA-----FVLADVIYTLIH----- 1474  
QY 118 WKSYCPRFPLDY-LEAFGLSDPLDH--QAVIKFEE--LETHSYVPSGVAP----- 166  
DB 1475 YINQRPSCIMDVLSRFSCLCDLISQVCQTAVYCKKALENHL--HVIYGTLLPLVYEQV 1532  
QY 167 ---HOYLSLLQDRYFPPIASVWRTLDKNFSLTPDLIHDLLGHV----- 206

DB 1533 EVQKQVLDL--KYL-----VIDKNQENNYITIKLDPFPHVFKDLRTQOKIKYSRG 1586  
QY 207 PMLHPSEFPIINMGRLEFTKVIKQVQALPSKKORIQTLSNLIAYRCFWFVSGGLE 266  
DB 1587 PSLLEBINH-FLSVSYDALPULTRBGLKDLRQLELHKQWDIMRASQDNPDGIMV 1645  
QY 267 -----NHEGRK---AYGAVL 278  
DB 1646 KLVVNLQLSKMAINHTEKEVLEAVGSCL 1675

RESULT 46  
US-08-952-127-3  
Sequence 3, Application US/08952127  
Patent No. 6211336  
GENERAL INFORMATION:  
APPLICANT: Shiloh, Yosef  
APPLICANT: Tagle, Danilo A.  
TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kohn & Associates  
STREET: 30500 No. 6211336thwestern Hwy., Suite 410  
CITY: Farmington Hills  
STATE: Michigan  
COUNTRY: U.S.  
ZIP: 48334  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/952,127  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Kohn, Kenneth I.  
REGISTRATION NUMBER: 30,995  
REFERENCE/DOCKET NUMBER: 2290,00029  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 810-539-5050  
TELEFAX: 810-539-5050  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3056 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-08-952-127-3

Query Match  
Best Local Similarity 4.3%; Score 81.5; DB 4; Length 3056;  
Pred. No. 47;  
Matches 61; Conservative 58; Mismatches 114; Indels 97; Gaps 17;

QY 10 PKYILKIAL-----KLRQSLSLFQNSQSLQRAVSPYVYRIIL--QKENKQ 57  
DB 1382 PSHVIAKATPAVYISNCHKTKLSILEILSK-----SPDSYQKILALICEQALETNN 1431  
QY 58 ALARHKCISILEFFKNLLEFVHLISLKNQREGCTDMAVSTPEFNRNLMYLLSRSFSL 117  
DB 1432 VYKGRILKIYHLFVSL-----KDKSGLGAWA-----FVLADVIYTLIH----- 1474  
QY 118 WKSYCPRFPLDY-LEAFGLSDPLDH--QAVIKFEE--LETHSYVPSGVAP----- 166  
DB 1475 YINQRPSCIMDVLSRFSCLCDLISQVCQTAVYCKKALENHL--HVIYGTLLPLVYEQV 1532  
QY 167 ---HOYLSLLQDRYFPPIASVWRTLDKNFSLTPDLIHDLLGHV----- 206

Db 1533 EVQKQVLDLL--KYL-----VIDNKNENLYITIKLLDPFPHVVFVKDLRITQOKIKYSG 1586  
 QY 207 PWLLHPSFSEFFINMGRLEFTKVIKQVQALPSKQRIOTLOSNIATVRCFWFTVESGLIE 266  
 Db 1587 PFSLEBINH-FLSVSVYDALPLTRLEGLKDLRQLEHLKQDMQVIMRASQDNPDQGIWV 1645  
 QY 267 -----NHEGRK-----AVGAVL 278  
 Db 1646 KLVNLLQLSKMAINTGEKVELEAVGSCL 1675

RESULT 47  
 US-08-952-014C-3  
 ; Sequence 3, Application US/08952014C  
 ; Patent No. 6265158  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Shilon, Yosef  
 ; TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE AND ITS  
 ; FILE REFERENCE: GENOMIC ORGANIZATION  
 ; NUMBER OF SEQUENCES: 91  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Kohn & Associates  
 ; STREET: 30500 No. 6265158thwestern Hwy., Suite 410  
 ; CITY: Farmington Hills  
 ; STATE: Michigan  
 ; COUNTRY: U.S.  
 ; ZIP: 48334  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/952,014C  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Kohn, Kenneth I.  
 ; REGISTRATION NUMBER: 30,995  
 ; REFERENCE/DOCKET NUMBER: 2290.00028  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 810-539-5050  
 ; TELEFAX: 810-539-5055  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; TYPE: amino acids  
 ; LENGTH: 3056 amino acids  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Homo sapiens

US-08-952-014C-3  
 Query Match 4.3%; Score 81.5; DB 4; Length 3056;  
 Best Local Similarity 18.5%; Pred. No. 47; Indels 97; Gaps 17;  
 Matches 61; Conservative 58; Mismatches 114;

QY 10 PKYLTKAL-----KLRQSLFFQNSQSORAYSTPYSYRIIL---OKENKEQ 57  
 Db 1382 PSHVikatFAYISNCHTKLKSILELSK-----SPDSYQKILLAIQAAETNN 1431  
 QY 58 ALARHKISILEFPKMLLFVHLLSLSKNORECSTDMVSTPFNNRLWYRLLSRFSL 117  
 Db 1432 VYKGRILKIYHLFVSLLL-----KDKSLGGAWA-----FVLRDVIYTLIH----- 1474  
 QY 118 WKSYPREFLDY-LEAFGLSLDLH--QAVIKFFE--LETHFSYYPVSGFVAP----- 166  
 Db 1475 YINORPSCIMDVLSRSLSCDLLSQVCQTAVTYCKDALENHL--HVIVGTLPVYEQV 1532  
 QY 167 ---HQYLSLQDRYFFPIASVMRTLDKDNFSLTPDLIHDLGHV----- 206

Db 1533 EVQKQVLDLL--KYL-----VIDNKNENLYITIKLLDPFPHVVFVKDLRITQOKIKYSG 1586  
 QY 207 PWLLHPSFSEFFINMGRLEFTKVIKQVQALPSKQRIOTLOSNIATVRCFWFTVESGLIE 266  
 Db 1587 PFSLEBINH-FLSVSVYDALPLTRLEGLKDLRQLEHLKQDMQVIMRASQDNPDQGIWV 1645  
 QY 267 -----NHEGRK-----AVGAVL 278  
 Db 1646 KLVNLLQLSKMAINTGEKVELEAVGSCL 1675

RESULT 48  
 US-09-360-416-3  
 ; Sequence 3, Application US/09360416  
 ; Patent No. 6438536  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Richard A. Gatti  
 ; TITLE OF INVENTION: METHODS FOR DETECTION OF ATAXIA  
 ; FILE REFERENCE: TELANGIECTASIA MUTATIONS  
 ; CURRENT APPLICATION NUMBER: US/09/360,416  
 ; CURRENT FILING DATE: 1999-07-23  
 ; NUMBER OF SEQ ID NOS: 143  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 3  
 ; LENGTH: 3057  
 ; TYPE: PRT  
 ; ORGANISM: Human

US-09-360-416-3  
 Query Match 4.3%; Score 81.5; DB 4; Length 3057;  
 Best Local Similarity 18.5%; Pred. No. 47;  
 Matches 61; Conservative 58; Mismatches 114; Indels 97; Gaps 17;

QY 10 PKYLTKAL-----KLRQSLFFQNSQSORAYSTPYSYRIIL---OKENKEQ 57  
 Db 1383 PSHVikatFAYISNCHTKLKSILELSK-----SPDSYQKILLAIQAAETNN 1432  
 QY 58 ALARHKISILEFPKMLLFVHLLSLSKNORECSTDMVSTPFNNRLWYRLLSRFSL 117  
 Db 1433 VYKGRILKIYHLFVSLLL-----KDKSLGGAWA-----FVLRDVIYTLIH----- 1475  
 QY 118 WKSYPREFLDY-LEAFGLSLDLH--QAVIKFFE--LETHFSYYPVSGFVAP----- 166  
 Db 1476 YINORPSCIMDVLSRSLSCDLLSQVCQTAVTYCKDALENHL--HVIVGTLPVYEQV 1533  
 QY 167 ---HQYLSLQDRYFFPIASVMRTLDKDNFSLTPDLIHDLGHV----- 206  
 Db 1534 EVQKQVLDLL--KYL-----VIDNKNENLYITIKLLDPFPHVVFVKDLRITQOKIKYSG 1587  
 QY 207 PWLLHPSFSEFFINMGRLEFTKVIKQVQALPSKQRIOTLOSNIATVRCFWFTVESGLIE 266  
 Db 1588 PFSLEBINH-FLSVSVYDALPLTRLEGLKDLRQLEHLKQDMQVIMRASQDNPDQGIWV 1646  
 QY 267 -----NHEGRK-----AVGAVL 278  
 Db 1647 KLVNLLQLSKMAINTGEKVELEAVGSCL 1676

RESULT 49  
 US-08-540-804-18  
 ; Sequence 18, Application US/08540804  
 ; Patent No. 5919666  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Young, Richard A.  
 ; APPLICANT: Koleske, Anthony J.  
 ; APPLICANT: Thompson, Craig M.  
 ; APPLICANT: Chao, David M.  
 ; TITLE OF INVENTION: No. 5919666el Factors Which Modify Gene  
 ; TITLE OF INVENTION: Transcription and Methods of Use Therefor  
 ; NUMBER OF SEQUENCES: 39  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.ra1

Page 26

STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/540,804  
FILING DATE: 11-OCT-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/521,872  
FILING DATE: 21-AUG-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,265  
FILING DATE: 25-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: WHI94-03A2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 323 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-540-804-18

Query Match 4.3%; Score 81; DB 2; Length 323;  
Best Local Similarity 23.0%; Pred. No. 1.7;  
Matches 56; Conservative 31; Mismatches 71; Indels 86; Gaps 12;  
QY 122 CRRFLDYI-EAFGLSDFL--DHQAV--IKFELETHFSYVSGFVAPHOYLSLQDR 176  
DB 134 CQYIRTLVSEARTLMPERFIPDPKYTEFEFYLLBELSY-----LIYVHPYQSLKQ-- 186  
QY 177 YFPILASVVR-----TLDKONFSLTPDLIHDLGHVWMLHPSFSEFFINMGRLETKVIE 230  
DB 187 ---IVQVLKQPPQITLSSDQLQNCWSLINDSYINDVHLLYPRH----- 227  
QY 231 KVOALPSEKKRIQTLQSNLAIYRCFWFTVESGLIENHGRKAYGAVLISSPQELGHAFI 290  
DB 228 -----IIVACLFTIT-----SIHGKPTKGSLSAASE----- 256  
QY 291 DNRVVLPLELDQIIRLPNTSTPOETLFSIRHPELVELTSKLEMMLDQGLIESIPLYNQ 350  
DB 257 -----AIRDPKNSSSPVOIAFNRMAESLVD---LEEVMDT-TQEQITLYDH 299  
QY 351 -EKY 353  
DB 300 WDKY 303

RESULT 50  
US-08-218-265-18  
Sequence 18, Application US/08218265  
Patent No. 5922585  
GENERAL INFORMATION:  
APPLICANT: Young, Richard A.  
APPLICANT: Thompson, Anthony J.  
APPLICANT: Koleske, Craig M.  
TITLE OF INVENTION: No. 5822585el Factors Which Modify Gene  
NUMBER OF SEQUENCES: 35  
TRANSCRIPTION AND METHODS OF USE THEREOF  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Militia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: US  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/218,265  
FILING DATE: 25-MAR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: WHI94-03  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 323 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-218-265-18

Query Match 4.3%; Score 81; DB 2; Length 323;  
Best Local Similarity 23.0%; Pred. No. 1.7;  
Matches 56; Conservative 31; Mismatches 71; Indels 86; Gaps 12;  
QY 122 CRRFLDYI-EAFGLSDFL--DHQAV--IKFELETHFSYVSGFVAPHOYLSLQDR 176  
DB 134 CQYIRTLVSEARTLMPERFIPDPKYTEFEFYLLBELSY-----LIYVHPYQSLKQ-- 186  
QY 177 YFPILASVVR-----TLDKONFSLTPDLIHDLGHVWMLHPSFSEFFINMGRLETKVIE 230  
DB 187 ---IVQVLKQPPQITLSSDQLQNCWSLINDSYINDVHLLYPRH----- 227  
QY 231 KVOALPSEKKRIQTLQSNLAIYRCFWFTVESGLIENHGRKAYGAVLISSPQELGHAFI 290  
DB 228 -----IIVACLFTIT-----SIHGKPTKGSLSAASE----- 256  
QY 291 DNRVVLPLELDQIIRLPNTSTPOETLFSIRHPELVELTSKLEMMLDQGLIESIPLYNQ 350  
DB 257 -----AIRDPKNSSSPVOIAFNRMAESLVD---LEEVMDT-TQEQITLYDH 299  
QY 351 -EKY 353  
DB 300 WDKY 303

Search completed: January 9, 2003, 17:00:47  
Job time : 42 secs



GenCore version 5.1.3  
Copyright (C) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2003, 14:44:28 ; Search time 81 Seconds  
(without alignments)  
920.853 Million cell updates/sec

Title: US-09-438-185A-1047

Perfect score: 1889

Sequence: 1 VHYCERTLPKYLKIALKL.....ESIPLYNQEKYLSGFVLQC 362

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

1: SPREMEL\_21.\*

2: sp\_archaea.\*

3: sp\_bacteria.\*

4: sp\_fungi.\*

5: sp\_human.\*

6: sp\_invertebrate.\*

7: sp\_mammal.\*

8: sp\_mhc.\*

9: sp\_organelle.\*

10: sp\_phage.\*

11: sp\_plant.\*

12: sp\_rodent.\*

13: sp\_virus.\*

14: sp\_unclassified.\*

15: sp\_rvirus.\*

16: sp\_bacteriap.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	237.5	12.6	491	13	Q9PU40
2	224.5	11.9	497	5	Q96370
3	219	11.6	579	5	Q24000
4	219	11.6	579	5	Q8SY95
5	215.5	11.4	453	11	Q91WV1
6	213.5	11.3	452	4	Q8TEY0
7	213.5	11.3	452	4	Q8TC14
8	206.5	10.9	244	2	Q9AG78
9	206.5	10.9	522	5	Q23438
10	206.5	10.9	532	5	Q9XZD1
11	181.5	9.6	457	5	Q9XYQ5
12	178.5	9.4	450	5	Q96947
13	178.5	9.4	495	2	Q9W0K2
14	177	9.4	495	2	Q9PDC3
15	157	8.3	438	5	Q17498
16	135.5	7.2	250	6	Q95LQ6

17	109.5	5.8	1253	10	Q9SCZ3	Q9scz3 arabidopsis
18	106.5	5.6	506	8	Q9GF63	Q9gf63 arabis alpi
19	106	5.6	129	13	Q42428	Q42428 lates calica
20	106	5.6	504	8	Q9GF55	Q9gf55 cardamine p
21	106	5.6	504	8	Q9GF31	Q9gf31 cardamine p
22	106	5.6	504	8	Q9GF30	Q9gf30 cardamine r
23	104	5.5	513	8	Q95F33	Q95f33 muhlenbergi
24	104	5.5	513	8	Q95F29	Q95f29 enneapogon
25	103.5	5.5	513	8	Q95F39	Q95f39 monanthochl
26	103.5	5.5	1782	5	Q9VP19	Q9vp19 drosophila
27	103	5.5	504	8	Q9GF41	Q9gf41 aubrieta de
28	103	5.5	513	8	Q9TI88	Q9tib8 bouteloua c
29	103	5.5	513	8	Q95F60	Q95f60 buchloe dac
30	103	5.5	513	8	Q95F50	Q95f50 schedonnard
31	103	5.5	513	8	Q95F38	Q95f38 reederchlo
32	102.5	5.4	1154	5	Q9VM23	Q9vm23 drosophila
33	102	5.4	506	8	Q9GF64	Q9gf64 arabis alpi
34	101.5	5.4	519	8	Q9MV23	Q9mv23 oryza meyer
35	101.5	5.4	1615	5	Q8T2D4	Q8t2d4 dictyosteli
36	101	5.3	635	17	Q972M5	Q972m5 sulfolobus
37	101	5.3	2042	5	Q9W1C5	Q9w1c5 drosophila
38	100.5	5.3	102	11	P97517	P97517 phodopus su
39	100.5	5.3	2212	12	Q9DD01	Q9dd01 ebola virus
40	99.5	5.3	513	8	Q9TI97	Q9ti97 sporobolus
41	99	5.2	513	8	Q95F61	Q95f61 agropogon c
42	99	5.2	513	8	Q95F62	Q95f62 symopogon
43	99	5.2	924	11	Q54321	Q54321 rattus norv
44	99	5.2	1078	17	Q8TFH6	Q8tfh6 methanosarc
45	98.5	5.2	513	8	Q9TIC0	Q9tico zeuglites pi
46	98.5	5.2	584	16	Q9PLZ8	Q9plz8 campylobact
47	98.5	5.2	945	10	Q94DZ7	Q94dz7 oryza sativ
48	98	5.2	924	11	Q9D4H1	Q9d4h1 mus musculu
49	98	5.2	1887	5	Q9GZG6	Q9gzg6 caenorhabdi
50	97.5	5.2	239	5	Q9NJQ3	Q9njq3 lymnaea sta
51	97.5	5.2	504	8	Q9GF36	Q9gf36 lepidium ca
52	97.5	5.2	519	8	Q9MV31	Q9mv31 oryza alta
53	97.5	5.2	519	8	Q9MV30	Q9mv30 oryza latif
54	97.5	5.2	519	8	Q9MV29	Q9mv29 oryza grand
55	97.5	5.2	946	10	Q94E00	Q94e00 oryza sativ
56	97	5.1	347	4	Q13786	Q13786 homo sapien
57	97	5.1	504	8	Q9GF53	Q9gf53 arabis proc
58	97	5.1	504	8	Q9GF28	Q9gf28 crucihimala
59	97	5.1	513	8	Q95F32	Q95f32 muhlenbergi
60	97	5.1	519	8	Q9MV25	Q9mv25 oryza ridle
61	97	5.1	519	8	Q9MV24	Q9mv24 oryza longi
62	96.5	5.1	363	8	Q9GHV8	Q9ghv8 caryodaphno
63	96	5.1	515	8	Q9GHD7	Q9ghd7 helionopsis
64	96	5.1	2059	12	Q56278	Q56278 human herpe
65	95.5	5.1	355	8	Q9GHT3	Q9ght3 mezilaurus
66	95.5	5.1	355	8	Q9GHM9	Q9ghm9 sassafras t
67	95.5	5.1	536	16	P96714	P96714 bacillus su
68	95.5	5.1	993	8	Q94QR6	Q94qr6 pleurotus o
69	95	5.1	2185	12	Q9E348	Q9e348 apple stem
70	95	5.0	502	8	Q9GF40	Q9gf40 arabis blep
71	95	5.0	504	8	Q9GF38	Q9gf38 crucihimala
72	95	5.0	970	10	Q22941	Q22941 arabidopsis
73	95	5.0	1147	5	Q9NSY5	Q9nsy5 caenorhabdi
74	94.5	5.0	245	12	Q91MT1	Q91mt1 lumpy skin
75	94.5	5.0	355	8	Q9GHM2	Q9ghm2 tambourissa
76	94.5	5.0	513	8	Q95F55	Q95f55 chloris tru
77	94.5	5.0	519	8	Q9MV37	Q9mv37 oryza offic
78	94.5	5.0	519	8	Q9MV36	Q9mv36 oryza eichi
79	94.5	5.0	519	8	Q9MV35	Q9mv35 oryza rhizo
80	94.5	5.0	699	16	Q9PND4	Q9pnd4 campylobact
81	94.5	5.0	212	12	Q91DD4	Q91dd4 ebola virus
82	94	5.0	293	17	Q9ZMC6	Q9zmc6 helicobacte
83	94	5.0	317	17	Q58067	Q58067 pyrococcus
84	94	5.0	504	8	Q9GF62	Q9gf62 barbarea vu
85	94	5.0	504	8	Q9GF51	Q9gf51 arabidopsis
86	94	5.0	504	8	Q9GF33	Q9gf33 microthlaspi
87	94	5.0	513	8	Q9TIA8	Q9tia8 perotis rar
88	94	5.0	513	8	Q95F36	Q95f36 tuctoria gr
89	94	5.0	618	16	Q9X0Q2	Q9x0q2 thermotoga

90	94	5.0	2079	12	09JGP2	09JGP2 apple latens
91	93.5	4.9	513	8	09TRI1	09TRI1 eustachys d
92	93.5	4.9	513	8	09SP35	09SP35 spartina pe
93	93	4.9	307	16	09CEX6	09CEX6 lactococcus
94	93	4.9	456	4	09BUC6	09BUC6 homo sapien
95	93	4.9	504	8	09GF94	09GF94 arabis hirs
96	93	4.9	504	8	09GF46	09GF46 halimolobos
97	93	4.9	513	8	09SF26	09SF26 tetrahene d
98	93	4.9	515	8	09GHC8	09GHC8 helionopsis
99	93	4.9	631	4	09G920	09G920 homo sapien
100	93	4.9	645	5	P91385	P91385 caenorhabdi
101	93	4.9	714	10	098RL8	098RL8 guillardi
102	93	4.9	1022	5	076546	076546 dictyostel
103	93	4.9	1301	16	08XNG4	08XNG4 clostridium
104	93	4.9	1391	5	017772	017772 caenorhabdi
105	92.5	4.9	433	11	08R2A0	08R2A0 mus musculu
106	92.5	4.9	513	8	09SF53	09SF53 enteropogon
107	92.5	4.9	519	8	09WV22	09WV22 oryza granu
108	92.5	4.9	735	3	09Y724	09Y724 schistosac
109	92.5	4.9	828	17	08UIB7	08UIB7 pyrococcus
110	92.5	4.9	1084	16	09PR18	09PR18 ureaplasma
111	92	4.9	355	8	09GHV6	09GHV6 hypodaphn
112	92	4.9	469	8	08SM78	08SM78 litsea umbe
113	92	4.9	502	8	08GP29	08GP29 silybrium
114	92	4.9	504	8	09G227	09G227 arabidopsis
115	92	4.9	504	8	09G1K9	09G1K9 arabidopsis
116	92	4.9	504	8	09GF61	09GF61 arabidopsis
117	92	4.9	504	8	09GF26	09GF26 ionopsidum
118	92	4.9	504	8	09GF25	09GF25 arabidopsis
119	92	4.9	504	8	09GF24	09GF24 arabidopsis
120	92	4.9	512	8	09G111	09G111 lillium conc
121	92	4.9	513	8	09T199	09T199 orcutia ca
122	92	4.9	513	8	09SF40	09SF40 vaseyochloa
123	92	4.9	515	8	09GHD3	09GHD3 helionopsis
124	92	4.9	515	8	09GHC9	09GHC9 helionopsis
125	92	4.9	620	11	008984	008984 ratuus novy
126	92	4.9	708	16	08RB41	08RB41 rathus novy
127	92	4.9	836	4	013785	013785 homo sapien
128	92	4.9	3262	4	013788	013788 homo sapien
129	91.5	4.8	355	8	09GH20	09GH20 chlorocardi
130	91.5	4.8	504	8	09GF57	09GF57 arabis pauc
131	91.5	4.8	504	8	09SBE6	09SBE6 gossypium l
132	91.5	4.8	513	8	09MUX5	09MUX5 orthoclada
133	91.5	4.8	513	8	09SFE7	09SFE7 chloris can
134	91.5	4.8	519	8	09WV34	09WV34 oryza punc
135	91.5	4.8	519	8	09WV32	09WV32 oryza punc
136	91.5	4.8	519	8	09WV28	09WV28 oryza minuc
137	91.5	4.8	519	8	09WV28	09WV28 oryza minuc
138	91.5	4.8	670	10	098RT4	098RT4 guillardi
139	91.5	4.8	907	10	09FW49	09FW49 arabidopsis
140	91	4.8	504	8	09GF52	09GF52 arabis puni
141	91	4.8	504	8	09GF45	09GF45 arabis turt
142	91	4.8	504	8	09GF27	09GF27 ionopsidum
143	91	4.8	513	8	09T192	09T192 fingerhutchi
144	91	4.8	513	8	09T191	09T191 uniola pani
145	91	4.8	513	8	09MXD0	09MXD0 digitaria s
146	91	4.8	513	8	09MUD3	09MUD3 sporobolus
147	91	4.8	515	8	09GHD6	09GHD6 helionopsis
148	91	4.8	1081	5	09VNU7	09VNU7 drosophila
149	91	4.8	1233	3	09PEL8	09PEL8 schistosac
150	91	4.8	2183	12	09WVY7	09WVY7 measles vir

## ALIGNMENTS

RESULT 1  
ID Q9PU40 PRELIMINARY; PRT; 491 AA.  
AC Q9PU40; (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE	Tyrosine hydroxylase.
OS	Gallus gallus (Chicken).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC	Gallus.
OK	NCBI_TaxID=9031;
RN	(1)
RP	SEQUENCE FROM N.A.
RL	Ernsberger U.;
RA	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN	[2]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=96076133; PubMed=7577670;
RA	Ernsberger U., Patzke H., Tiseler-Seta J.P., Reh T., Goridis C.,
RA	Rohrer H.;
RT	"The expression of tyrosine hydroxylase and the transcription factors
RT	cPhox-2 and Cash-1: evidence for distinct inductive steps in the
RT	differentiation of chick sympathetic precursor cells."
RL	Mech. Dev. 52:125-136(1995).
DR	EMBL; AJ251387; CAB62388.1; -
DR	HSSP; P04177; 1T0H.
DR	InterPro; IPR001273; Aaa_hydroxylase.
DR	Pfam; PF00351; bioprotein_H_1.
DR	PRINTS; PR00372; PYMHYDRXLASE.
DR	TIGRPFMS; TIGR01269; TYR3_MONOOXYL_1.
DR	PROSITE; P500367; BIOPROTEIN_HYDROXYL_1.
DR	SEQUENCE 491 AA; 56016 MW; 6322F2D38E746930 CRC64;
SO	SEQUENCE
Query Match	12.6%; Score 237.5; DB 13; Length 491;
Best Local Similarity	29.1%; Pred. No. 42e-12;
Matches	66; Conservative 36; Mismatches 98; Indels 27; Gaps 4;
QY	107 WYRLSSRFSLMYSYCPREFLDYLAFLGSLSPDLH-----QAVKFELETFPSY 158
DB	226 WKEVYSLTKSLVYTHACK--EYLAFLMLEKFCGYMENNIPOLIEFSRLKERTFOUR 282
QY	159 PVSGFVPHOYLILDRYFPPLASVWRTLDKDNFSLPDLIHDLGHVPLHPSSEFP 218
DB	283 PVAGLLSARDFLASLAFVFOCTQYIRHASSPMHSPEPDCCHLGHVPLADKTPAQS 342
QY	219 INMRRLPTKYEVQALPKKORIOTLOSNIATYRCWFWVESGLLENHGKAYAVY 278
DB	343 QDIG-----LASLGATDEIEKATL-----YMFVTEGLCRONGIKYAKAGGL 386
QY	279 ISSPQELGAFIDNVAVLPLELDQIRLPENTSTPOETLPSIRHDE 325
DB	387 LSTIGELHSLSDPEVYRDPDPAANVPIDQNVQPVYFVSESFS 433

RESULT 2  
ID Q96370 PRELIMINARY; PRT; 497 AA.  
AC Q96370; (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Tryptophan hydroxylase (EC 1.14.16.4).  
GN TPH.  
OS Schistosoma mansoni (Blood fluke).  
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidae;  
OC Schistosomatidae; Schistosomatidae; Schistosoma.  
OX NCBI\_TaxID=6183;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PUERTO RICAN;  
RX MEDLINE=99348305; PubMed=10419486;  
RA Hamdan F.P.; Ribeiro P.P.;  
RT "Characterization of a stable form of tryptophan hydroxylase from the  
human parasite Schistosoma mansoni."  
RL U. Biol. Chem. 274:21746-21754(1999).  
DR EMBL; AF031034; AAD01923.1; -  
DR HSSP; P04177; 1T0H.  
DR InterPro; IPR001273; Aaa\_hydroxylase.



OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OK NCBI\_TaxID=7227;  
 RN (1)  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RA Stagleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champagne M., Chavez C., Dorsett V., Dresnek D., Fartan D., Fisse E.,  
 RA George R., Gonzalez M., Guatin H., Kronmiller B., Li P., Lao G.,  
 RA Miranda A., Mungall C.J., Nunco J., Pacleb J., Paragas V., Park S.,  
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
 RA Celisner S.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY071698; AAL49320.1; -  
 SQ SEQUENCE 579 AA; 66012 MW; 7700F94E436479E2 CRC64;

Query Match 11.6%; Score 219; DB 5; Length 579;  
 Best Local Similarity 28.6%; Pred. No. 28-10; Mismatches 94; Indels 54; Gaps 8;  
 Matches 74; Conservative 37; Mismatches 94; Indels 54; Gaps 8;

QY 128 DYLEAFGLSD---FLDH---QAVIKFPELTHPSYVSGFVAPHOYSLIDRYFP 179  
 DB 328 EVRAHFOKLODEQIFVERRLPQLOEMSDFLAKRTGSLAPAGLTLADFLASLAFIQ 388  
 QY 180 IASVWRITDKDNFSLTPDLIHDLGVPMILHPSSEFPIINGRFLTXYIEKVALPSKK 239  
 DB 388 STQYRVHVSFPHTEPSPSHLGHMPLADSPFQSOEIG-----LASLGASD 439  
 QY 240 QRIQTLQSLIAIVRCFMTVESGLIENHGRKAYGAVLISSPOELGAFID- 291  
 DB 440 ESEIKLST-----VYMFVEFGLCCKEHOQIAYAGALLSYGLLHAISDKENHRAFE 492  
 QY 292 -NVRVPLE-----LDQIRLPNTSTPQETLFSIRHPD-ELVELT 330  
 DB 493 PASTVAVQYQOEYQPIYYAASFEDAKDKFRRWSTMRPFEVRFN-PHTERVEVLDSV 551  
 QY 331 SKLEMMLDQGLLESIPLYN 349  
 DB 552 DKLGLTVHQMNTTEILHLTN 570

RESULT 5  
 O91WV1 PRELIMINARY; PRT; 453 AA.  
 AC O91WV1.  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Phenylalanine hydroxylase.  
 OS Mus musculus (mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OK NCBI\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=KIDNEY;  
 RC Strausberg R.;  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC013458; AAI13458.1; -  
 DR InterPro; IPR001273; Aaa\_hydroxylase.  
 DR InterPro; IPR002912; ACT.  
 DR Pfam; PF01842; ACT; 1.  
 DR PRINTS; PRO0351; bioprotein H; 1.  
 DR TIGRPFAM; TIGR01268; Phehydroxylase.  
 DR PROSITE; PS00367; BIOTRIN HYDROXYL; UNKNOWN 1.  
 SQ SEQUENCE 453 AA; 51899 MW; 551P181FA59DA5B CRC64;

Query Match 11.4%; Score 215.5; DB 11; Length 453;  
 Best Local Similarity 27.8%; Pred. No. 2.9e-10;  
 Matches 64; Conservative 40; Mismatches 105; Indels 21; Gaps 5;

QY 104 RNLM---YLLSSRFSLMSYCPREFLDYLEAF-GLSDFLDH-QAVIKFPELTHPSY 158  
 DB 184 KRTWGTIVFKTKLVKTHACYEHNHIFPLELEKCFREDNIPQLEDVSOQLQCTGRUR 243  
 QY 159 PVSGFVAPHOYSLIDRYFPPIASVWRITDKDNFSLTPDLIHDLGVPMILHPSSEFF 218  
 DB 244 PVAGLSSRFGLGLAFRVFHCOTYIRHSGSKMPTPEPDICHELHGVPLPSDRSFAQS 303  
 QY 219 INMGRLFTXYIEKVALPSKKORIQTLOSNIATVRCFMTVESGLIENHGRKAYGAVL 278  
 DB 304 QEID-----LASLGAPDEYIEKLTAT-----IYMFVEFGLCCKEDSIRKAYGAGL 347  
 QY 279 ISSPOELGAFIDNVAVLPLELDQIRLPNTSTPQETLFSIRFDELVE 328  
 DB 348 LSSFGLQYCLSDKPKLPLPLELEKTAQCEYTVTERQPLYYVASEFNDAKE 397

RESULT 6  
 O8TEYO PRELIMINARY; PRT; 452 AA.  
 AC O8TEYO.  
 DT 01-JUN-2002 (Tremblrel. 21, Created)  
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Phenylalanine hydroxylase.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OK NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Konecki D.S., Licher-Konecki U.;  
 RT "Completion of the sequence of PAH, a model disease gene."  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF404777; AAL78816.1; -  
 SQ SEQUENCE 452 AA; 51861 MW; DEF9DB9D6B8C800C CRC64;

Query Match 11.3%; Score 213.5; DB 4; Length 452;  
 Best Local Similarity 26.6%; Pred. No. 4.4e-10;  
 Matches 62; Conservative 44; Mismatches 100; Indels 27; Gaps 4;

QY 104 RNLM---YLLSSRFSLMSYCPREFLDYLEAFGLSDFLDH-----QAVIKFPELTHP 155  
 DB 184 KRTWGTIVFKTKLVKTHA---CYEYNIIFPLELEKCGFREDNIPQLEDVSOQLQCTGR 240  
 QY 156 SYVSGFVAPHOYSLIDRYFPPIASVWRITDKDNFSLTPDLIHDLGVPMILHPSFS 215  
 DB 241 RLRVAGLSSRFGLGLAFRVFHCOTYIRHSGSKMPTPEPDICHELHGVPLPSDRSFA 300  
 QY 216 EPIINMGRLFTXYIEKVALPSKKORIQTLOSNIATVRCFMTVESGLIENHGRKAYG 275  
 DB 301 QFSQEID-----LASLGAPDEYIEKLTAT-----IYMFVEFGLCCKEDSIRKAYG 344  
 QY 276 AVISSPOELGAFIDNVAVLPLELDQIRLPNTSTPQETLFSIRFDELVE 328  
 DB 345 AGLSSFGELQYCLSEKPKLPLPLELEKTAQIONVTVTERQPLYYVASEFNDAKE 397

RESULT 7  
 O8TC14 PRELIMINARY; PRT; 452 AA.  
 AC O8TC14.  
 DT 01-JUN-2002 (Tremblrel. 21, Created)  
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Phenylalanine hydroxylase.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OK NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.

```

Db      235 DALYTDG 241
: : : |
RESULT 9
Q23438 PRELIMINARY; PRT; 522 AA.
ID Q23438
AC Q23438;
DT 01-NOV-1996 (TReMBUrel. 01, Created)
DT 01-OCT-2001 (TReMBUrel. 18, Last sequence update)
DT 01-JUN-2002 (TReMBUrel. 21, Last annotation update)
DE Hypothetical 59.8 kDa protein.
DE ZK1290.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OC NCBI_TaxID=6239;
OC [1]
RN RP SEQUENCE FROM N.A.
RN RP STRAIN=BRISTOL N2;
RN RX MEDLINE=99069613; PubMed=9851916;
RN None.
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.;"
RT Science 282:2012-2018 (1998).
RN [2]
RN RP SEQUENCE FROM N.A.
RN RP STRAIN=BRISTOL N2;
RN RC STRAIN=BRISTOL N2;
RN RA "atch A.;"
RN RT "The sequence of C. elegans cosmid ZK1290.;"
RN RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RN RP SEQUENCE FROM N.A.
RN RP STRAIN=BRISTOL N2;
RN RC STRAIN=BRISTOL N2;
RN RA "Direct Submission.;"
RN RT Waterston R.;"
RN RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN EMBL; U21308; AAB93319.2; -.
RN HSSP; P04176; 1PHZ.
RN InterPro; IPR001273; Aaa_hydroxylase.
RN Pfam; PF00351; biopetrin_H; 1.
RN PRINTS; PR00372; FYWHYDRXLASE.
RN TIGRFAMs; TIGR01270; Trp_5_monoox; 1.
KW Hypothetical protein.
SQ SEQUENCE 522 AA; 59781 MW; B6205C4B932C21FA CRC64;

Query Match 10.9%; Score 206.5; DB 5; Length 522;
Best Local Similarity 25.2%; Pred. No. 2.1e-09;
Matches 70; Conservative 47; Mismatches 116; Indels 45; Gaps

QY 104 RNLLW---YRLSSRFSLSWKSVCPRFFLDYLAFAGLLSDFLDH-----QAVIKPFPELETHF 155
Db 241 RKTWGIYRKLR--ELHKHKACKQFLDNFELLERHCGYSENNIPOLEDICKFLKAKTGF 297
QY 156 SYFVSGFVAPHQYLSLDQRYFPPIASWRTLDKONFSLTDLHLLGHVPWLLHSPES 215
Db 298 RVRPVAGLSLARDFLAGLAVRVFCTQYVRHADPFPTPEPDVVELHGMHALFADPDFA 357
QY 216 EFFINNGRLFTVKIEKQVALPSKKQRIQTLSNLAIIVRCFTWTVESGLI----- 265
Db 358 QFQSEIG-----LASGAGEEDLKKIATL-----YFSEIEFGSSDDAADSFKV 401
QY 266 ---ENHGRKAYGAVLISSPQELGHAFIDNVRVLPLELDQIRLPNTSTPQETLFSIRH 322
Db 402 ENGNSHERPKVYGAGLLSSAGELQHAVESATIRFQPDVRVWQEBCLITTFQSAIFYTRN 461
QY 323 FDELVELTSKLEWMLDQGLLESIPLYNQEKYLSGFEVL 360
Db 462 FEEAQQ---KLRFMTNNMKRPFIVRYN--PYTESVEVL 494

RESULT 10
O99XZD1

```

```

ID Q9XZD1 PRELIMINARY; PRT; 532 AA.
AC Q9XZD1;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE tryptophan hydroxylase.
GN TPH-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Size J.Y., Ruvkun G.;
RT "tpH-1 encodes a C. elegans tryptophan hydroxylase.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF135186; AAD30115.1; -.
DR HSSP; P04176; 1PH2
DR InterPro; IPR001273; Aaa_hydroxylase.
DR Pfam; PF00351; bioperlin_H.1.
DR PRINTS; PR00372; FYMHYDRXLASE.
DR TIGRfams; TIGR01270; tpd_5_monoox; 1.
SQ SEQUENCE 532 AA; 60863 MW; 8AC6A51C7DD0121F CRC64;

Query Match 10.9%; Score 206.5; DB 5; Length 532;
Best Local Similarity 25.2%; Pred. No. 2.1e-09;
Matches 70; Conservative 47; Mismatches 116; Indels 45; Gaps 8;

QY 104 RNLM---YRLSSRFSLWKSYPFLDYLEAFGLSDFLDH-----QAVIKFPELETHF 155
Db 251 RTWGIITIRKLR---ELHKHACKQFLDNFELLEHRHCGSENNIPOLBIDICFKLAKTGF 307
QY 156 SYYPVSGFVAPHOYLSLQDRYFPFIASVMRTLDKNFSLTPDLIDLGHWYWLHPSPFS 215
Db 308 RVRPAGYLSADFLAGLAYVVFCTQYVRHNDPPTYPEBPVHELMGHALFADPDFA 367
QY 216 EFFIMNGRLFTVIEKVQALPSKKORIQTQSNLAIIVACFMFTVSGLI----- 265
Db 368 QSGSEIG-----LASIGASEEDLKIATL-----YFSTIEFGLSDDAADSPVK 411
QY 266 ---ENHEGRKAYGAVLSSPOELGAFIDNVRVLPLEDOIIRLPNTSTQETLFSIRH 322
Db 412 ENGSHRFRKYGAGLSSAGELQHAVESATIRFDPDRAVQECILITFQSAVFTIRN 471
QY 323 FDELVELTSKLEWMLDGLLESIPLYNQEKYLSGFVYL 360
Db 472 FEEAQQ---KLRFMTNNMKRPFIIVRN--PYTESVEVL 504

RESULT 11
ID Q9XYO5 PRELIMINARY; PRT; 457 AA.
AC Q9XYO5;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Phenylalanine hydroxylase (EC 1.14.16.1).
GN KO8F8.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Loefer C.M., Davidson B., McKerrow J.;
RT "A phenylalanine hydroxylase gene from the nematode Caenorhabditis
RT elegans is expressed in the hypodermis.";
RL J. Neurogenet. 0:0-0(1999).
DR EMBL; AF119388; AAD31643.1; -.
DR HSSP; P00439; 4PAH.
DR InterPro; IPR001273; Aaa_hydroxylase.
DR TIGRfams; TIGR01273; Aaa_hydroxylase.
DR InterPro; IPR001273; Aaa_hydroxylase.
SQ SEQUENCE 457 AA; 52189 MW; 7A573B84B9BEF6FC CRC64;

```

```

DR InterPro; IPR000179; Cyt_b_b6.
DR Pfam; PF01842; ACT; 1.
DR Pfam; PF00351; bioperlin_H.1.
DR PRINTS; PR00372; FYMHYDRXLASE.
DR TIGRfams; TIGR01268; Phe4hydrox tetra; 1.
DR PROSITE; PS00367; BIOPTERIN HYDROXYL; 1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; UNKNOWN_1.
KW Oxidoreductase.
SQ SEQUENCE 457 AA; 52189 MW; 7A573B84B9BEF6FC CRC64;

Query Match 9.6%; Score 181.5; DB 5; Length 457;
Best Local Similarity 21.8%; Pred. No. 2.4e-07;
Matches 83; Conservative 66; Mismatches 137; Indels 95; Gaps 15;

QY 14 LKIALKROSLSPFONSQSLQRAY----STPYSYRIILQKENEKQALARKCISILE 69
Db 81 LKTIYRLKELSL---SNKKLRRLFKTGIPK-----KQKDSVPWFQKINDIDQ 130
QY 70 FFKNLL-----FVHLISLKNQEGCSTDMA-----VSTPFENRL--WYRL 110
Db 131 FANRLSYGALDADHPGFKDMTYRERKFPADIAFPKQGDKIPIITYDEBIATWRTV 190
QY 111 LSSRFSLW-KSYCPREFLDYLEAFGLSD-----FLDHQAVIKFPELETHFSYYPVS 161
Db 191 YNELTWVPYPMKACQEF--NYI--FPLQONCGFGRDIPOLQDVSFLKDCGTGTYTRPA 246
QY 162 GFVAPHOYLSLQDRYFPFIASVMRTLDKNFSLTPDLIDLGHWYWLHPSPSEPFIM 221
Db 247 GLSPRDFLAGMAFRVHFSTQYIRHNSAPKTYTPEDICHELGLHVPFADVPEAPQSGRI 306
QY 222 GRLE---TVIEKVQALPSKKORIQTQSNLAIIVACFMFTVSGLIENHEGRKAYGAV 277
Db 307 GLASLAPDQVIEIKIATL-----VFTIEFGICQDDGKRAYGAG 346
QY 278 LSSPOELGAFIDNVRVPLR--LDQIRLPNTSTQETLFSIRH 317
Db 347 LLSFGLQYALSDKREVDPDPAVCCTKPIPIEYQPKFLAESPAKAKLKSWAAT 406
QY 318 ---FSIRH--FDELVELTSKL 333
Db 407 NRPFQIRNAYVYQVAILDKV 427

RESULT 12
ID Q96947 PRELIMINARY; PRT; 450 AA.
AC Q96947;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Phenylalanine hydroxylase.
OS Geodia cydonium (Sponge).
OC Eukaryota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;
OC Astrophorida; Geodiidae; Geodia.
OX NCBI_TaxID=6047;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99093257; PubMed=9877430.
RT "Phenylalanine hydroxylase from the sponge Geodia cydonium:
RT Implication for allotrecognition and evolution of aromatic amino acids
RT hydroxylases.";
RL Dev. Comp. Immunol. 22:469-478(1998).
DR EMBL; Y16353; CAA76184.1; -.
DR HSSP; P00439; 2PAH.
DR InterPro; IPR001273; Aaa_hydroxylase.
DR InterPro; IPR002912; ACT.
DR Pfam; PF01842; ACT; 1.
DR Pfam; PF00351; bioperlin_H.1.
DR PRINTS; PR00372; FYMHYDRXLASE.
DR TIGRfams; TIGR01268; Phe4hydrox tetra; 1.
DR TIGRfams; TIGR01268; Phe4hydrox tetra; 1.
SQ SEQUENCE 450 AA; 51204 MW; 691880218BB9725D CRC64;

```

Query Match 9.4%; Score 178.5; DB 5; Length 450;  
Best Local Similarity 25.8%; Pred. No. 4.3e-07;  
Matches 55; Conservative 43; Mismatches 80; Indels 35; Gaps 6;  
QY 107 WYRLSSRFLWSKSYCPFRFLDYLEAFGLSDFLDH-----QAVIKFELETHFSY 158  
DB 186 WRTPTNLVLFPTHACK--EHNHVPFLQENCGYREDNIPQLEVSQYLSQCTGFLRL 242  
QY 159 PVSGFVAPHOYLSLQDRYFPPIASVMRTLDKDNFSLTDLHDHGLGHVPLLHPSPSEFF 218  
DB 243 PVAGLLSRDFLAGLAFRVHSHVQYIRHYSONYPTPEDVCHELIGHVRCV-ILFAQFS 301  
QY 219 INMGRLTKVIEKQVALPSKKQRIQTLQSNLIAIVRCFWFTVESGLIENHEGRKAYGAVL 278  
DB 302 QEIG-----LASGAPEEVVOGLAT-----LWFTIEFGLCKQDQGTQKAYGAGL 345  
QY 279 ISSQELGHAFINVRVLPLELDQIIRLPFNTS 311  
DB 346 ISSFGLQYCLSDKPEVRPLD-----PFKTS 371

RESULT 13  
Q9W0K2 PRELIMINARY; PRT; 555 AA.  
AC Q9W0K2;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE CG9122 protein.  
GN CG9122.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKLEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt R.G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA de Pablo B., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelist C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.B., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hoston D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lim P.P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Malshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusekern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Relnett K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Sidenkiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weisenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster";  
RL Science 287:2185-2195(2000).  
DR EMBL: AE003470; AAF47444.1; -;  
DR HSSP: P04176; 1PHZ.  
DR PLYbase: FB00035187; CG9122.  
DR InterPro: IPR001273; Aaa\_hydroxylase.  
DR InterPro: IPR002912; ACT.  
DR Pfam: PF01842; ACT.  
DR Pfam: PF00351; biotpterin H; 1.  
DR PRINTS: PR00372; FWHYDRXLAASE.  
DR TIGRFAMs: TIGR01270; TTP\_5\_monoox; 1.  
DR PROSITE: PS00367; BIOTPTERIN HYDROXYL; 1.  
SQ SEQUENCE 555 AA; 61530 MW; 60D527FEB1FA791C CRC64;

Query Match 9.4%; Score 178.5; DB 5; Length 555;  
Best Local Similarity 32.0%; Pred. No. 5.6e-07;  
Matches 47; Conservative 28; Mismatches 55; Indels 17; Gaps 3;

QY 143 QAVIKFELETHFSYYPVSGFVAPHOYLSLQDRYFPPIASVMRTLDKDNFSLTDLHDH 202  
DB 284 QDVSVYLKRKTGFLRPVAGYLSRDLGLAFRVHSHVQYIRHSSDPFYTPEDCCHEL 343  
QY 203 LGHVPLLHPSPSEFFINMGRFLTQVIEKQVALPSKKQRIQTLQSNLIAIVRCFWFTVES 262  
DB 344 LGHNPLLANSFPAQFSQEI-----LASGASDADIEKLATL-----YFTTVEF 387  
QY 263 GLIENHEGR-KAYGAVLISQPOELGHA 288  
DB 388 GLCKQADSTFKVVGAGLLSSVAELOHA 414

RESULT 14  
Q9FDC3 PRELIMINARY; PRT; 495 AA.  
AC Q9FDC3;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE PAH.  
GN PAH.  
OS Myxococcus xanthus.  
OC Bacteria; Proteobacteria; Delta subdivision; Myxobacteria;  
OC Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.  
OX NCBI\_TaxID=34;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21359321; PubMed=11466282;  
RA Sun H., Shi W.;  
RT "Genetic Studies of mrp, a Locus Essential for Cellular Aggregation  
and Sporulation of Myxococcus xanthus.";  
RL J. Bacteriol. 183:4786-4795(2001).  
DR EMBL: AF285263; AAR99527.1; -;  
DR InterPro: IPR001273; Aaa\_hydroxylase.  
DR Pfam: PF00351; biotpterin H; 1.  
SQ SEQUENCE 495 AA; 53213 MW; CD82B52F8A613AEB CRC64;

Query Match 9.4%; Score 177; DB 2; Length 495;  
Best Local Similarity 23.8%; Pred. No. 6.5e-07;  
Matches 69; Conservative 47; Mismatches 128; Indels 46; Gaps 6;

QY 91 STDMAVSTPFFNRNLWYRLSSRFLWSKSYCPFRFLDYLEAFGLSDFLDHQAQVIFPE 150  
DB 31 SRDAQV-----WRNIIGLSLRGLADKAHPVLEGLEATGIGSECTPSLDENNEKL 80  
QY 151 LETHFSYYPVSGFVAPHOYLSLQDRYFPPIASVMRTLDKDNFSLTDLHDHGLGHVPLL 210  
DB 81 ARLCMACVGVGRFIPPAVFTELQAMGVLAIAADIRTHEIETYPADPDIHVESAGAPIIA 140  
QY 211 HPSESEFFINMGRFLTQVIEK--GAL-----PSKKQRIQTLQSNLIA--- 251  
DB 141 NRRVAEYLCACGLGVGFKATSVEDQAVFATRNLSVVKEDPDASEEAAHAQARLEASA 200

QY 252 -----IVRCFWFTVESGLIENHEGRKAYGLISSPOLGHAFLDNVRLPLED 301  
 DB 201 SRRVSESTRASRLVMTAEVGLGIVASPRIVAGLFTSIEBACHLTPLVXKLPLSV- 259  
 QY 302 QIRLPNTSTPQETLSIRHPELVELTSKLEWMT-----DGLLESTI 345  
 DB 260 ACADMDYDITRMQPOLFVARDPEHLFEVLAEFESTLSMKRGDPGLTEAL 309

## RESULT 15

ID 017498 PRELIMINARY; PRT; 438 AA.  
 AC 017498;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)  
 DE Phenylalanine hydroxylase (EC 1.14.16.1) (Fragment).  
 GN PAH.  
 OS Branchiostoma floridae (Florida lancelet) (Amphioxus).  
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;  
 CC Branchiostoma.  
 OX NCBI\_TaxID=7739;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Paton S.J., Luke G.N., Holland P.W.H.;  
 RT "Complex history of a chromosomal paralog region: insights from  
 RT amphioxus aromatic amino acid hydroxylase genes and insulin-related  
 RT genes.";  
 RL Mol. Biol. Evol. 15:1373-1380 (1998).  
 DR EMBL; AB001677; CAA04917.1; -.  
 DR HSSP; P04176; 1PHZ.  
 DR InterPro; IPR001273; Aaa\_hydroxylase.  
 DR InterPro; IPR002912; ACT.  
 DR InterPro; IPR002119; Histone\_H2A.  
 DR Pfam; PF01842; ACT; 1.  
 DR Pfam; PF00351; bioprotein\_H; 1.  
 DR TIGRFAMs; TIGR01268; Phehydrox\_tetr; 1.  
 DR PROSITE; PS00046; HISTONE\_H2A; UNKNOWN\_1.  
 KW Oxidoreductase.  
 FT NON\_TER  
 SQ SEQUENCE 438 AA; 4958 MW; 1619297DBDBF5EE7 CRC64;

Query Match 8.3%; Score 157; DB 5; Length 438;  
 Best Local Similarity 24.1%; Pred. No. 2.9e-05;  
 Matches 64; Conservative 37; Mismatches 123; Indels 36; Gaps 8;

QY 105 NIWYRLSSRSFSLMKSYCPR----FIDYIEAFGLSDPLDH-QAVIKFELETHFSYYP 159  
 DB 170 NIWQITFELRLKLTPTDACREHNVFPLMENGCFREDNIPOLEDSNFKDCTGFTLRP 229  
 QY 160 VSGFVAHQVYSLLDQRYFPPLASVMTLDKONFSLTPDLIHDLGHVPLMLHPSPSEFFI 219  
 DB 230 VAGLLSRDPLAGLAPFVHFSTOYIRHNSKPLTTPPDVCHELGHAPLPADPSFAQPSQ 289  
 QY 220 NMGRLTVKIEQVALPSKCKORIGTOSNIIAIVRCFWFTVESGLIENHEGRKAYGAVLI 279  
 DB 290 ELG-----LASTGAPDDFWKATL-----YMTVFGACRODGEKAYGRDCC 333  
 QY 280 SSPQELGAFIDNVRLPLELDQIRLPNTSTPQETLSIRHPELVELTSKLEWMTDQ 339  
 DB 334 HNSESCRYCLTDKPDIRPFEBEKISVTKYPTTEYQPIYVADSFODAKEKVR--TWS--- 388  
 QY 340 GLLESIFL-----YNOEKYLSGFEVL 360  
 DB 389 ---HSIRPRFSVHN--PYQGSVEL 409

## RESULT 16

ID 095106 PRELIMINARY; PRT; 250 AA.  
 AC 095106;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)

DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)  
 DE Tyrosine hydroxylase (EC 1.14.16.2) (Fragment).  
 GN TH.  
 OS Equus caballus (Horse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
 OX NCBI\_TaxID=9796;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=THOROUGHRED; TISSUE=ADRENAL MEDULLA;  
 RA Sato F., Matsuta M., Hasegawa T.;  
 RT "Molecular cloning of mRNA for equine tyrosine hydroxylase."  
 RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AB071421; BAB68123.1; -.  
 DR InterPro; IPR001273; Aaa\_hydroxylase.  
 DR Pfam; PF00351; bioprotein\_H; 1.  
 DR PROSITE; PS00367; BIOPROTEIN\_HYDROXYL; UNKNOWN\_1.  
 KW Oxidoreductase.  
 FT NON\_TER  
 FT NON\_TER  
 SQ SEQUENCE 250 AA; 28548 MW; 293BCE9D98F39A6B CRC64;

Query Match 7.2%; Score 135.5; DB 6; Length 250;  
 Best Local Similarity 28.2%; Pred. No. 0.001;  
 Matches 35; Conservative 25; Mismatches 53; Indels 11; Gaps 2;

QY 107 WYRLSSRSFSLMKSYCPRFIDYIEAFGLSDPLDH-----QAVIKFELETHFSYYP 158  
 DB 124 WKEVYTTKGLVYTHACR--EHLFAFELERFSGYREDNIPOLEDSRFLKERSGFQUR 180  
 QY 159 VSGFVAHQVYSLLDQRYFPPLASVMTLDKONFSLTPDLIHDLGHVPLMLHPSPSEFF 218  
 DB 181 PVAGLSARDFLASIAFRVQCTOYIRHASSPMHSPEDDCHELGHVPLMDRTFAOPS 240  
 QY 219 INMG 222  
 DB 241 QDIG 244

## RESULT 17

ID 098CZ3 PRELIMINARY; PRT; 1253 AA.

AC 098CZ3;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)  
 DE Disease resistance-like protein.  
 GN F26013.200.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=5702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Delaney M., Berger C., Cooke R., Grellet F., Landie M., Mewes H.W.,  
 RA Lemcke K., Mayer K.F.X., Quetler F., Salanoubat M.;  
 RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Arabidopsis sequencing project;  
 RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AL133452; CAB63020.1; -.  
 DR InterPro; IPR000767; Disease\_resist.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR002182; NB-ARC.  
 DR InterPro; IPR00157; TIR\_domain.  
 DR Pfam; PF00560; LRR; C.  
 DR Pfam; PF00560; LRR; C.  
 DR Pfam; PF00561; NB-ARC; 1.  
 DR PRINTS; PRO0364; DISEASERSIST.  
 DR SMART; SM00255; TIR; 1.



```

SQ SEQUENCE 1253 AA; 140809 MW; 4211ADE056C2B01 CRC64;
Query Match 5.8%; Score 109.5; DB 10; Length 1253;
Best Local Similarity 20.3%; Pred. No. 1.2; Mismatches 130; Indels 127; Gaps 18;
Matches 81; Conservative 61;

Qy 77 VHLISLKNQRECS-----TMAVVSTPFFNRNLWYRLSS-----RF 115
Db 421 IHL-----EGCGFFPRVEINLVKCLVSA-EGVVVHNLQISGRKINGKRKS 472

Qy 116 SLKSYCPFFLDYLEAFG---LLSDFLDHOAV-----IKFFEL-----ET 153
Db 473 RLWKPLIYKFDRLQVLSGSEIDIAIFLDPSALSFDVNPMAFENMYNLRYLKICSSNPGN 532

Qy 154 HFSYVPVGFVAPHQVLSLQDRYPIASVMRTLDKDNFSL-----TPD-- 197
Db 533 HYALHLPKGVKSLPEELRLHWEHPFLSLPQDFTNRLNLVILNMCYSKLQRLWEGTKELG 592

Qy 198 -LIHDLHGVPMLLHPSFSEPPIN-----CRL-----FTKVIB-----KVQ 233
Db 593 MLKRIMLCHSQQLVGIQELQIALNWEVIDLQGCARLQRFATGCHQHLRVINLSGCIKIK 652

Qy 234 ALPSKKORIOTL---OSNLIAIVRCFWFTVESGLIENHEGRKAYGAVGLISSPOELG-HAF 289
Db 653 SFPEVPNTIELYKGTGIRSIPTVTFSPQNSFIYDKDKHFLNREVSSDSQSLSIWY 712

Qy 290 IDNRVRLP---LELDQIIRLPNT-----STQETLFSIRHPDELVEL-----TSKLEWM 336
Db 713 LDNLKVLDSLSCQLELEDIOGIPRNRLKYLGGTAIKELPSMLHLSLVLDLENCKRLHK 772

Qy 337 LDQGL-----LESIPLYNQEKYLSG 356
Db 773 LPMGIGNLSLAVINLSGCGSELEDIOGIPRNLEELYLAG 811

RESULT 18
ID Q9GF63 PRELIMINARY; PRT; 506 AA.
AC Q9GF63;
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Probable intron maturase (Maturase K).
GN MATK
OS Arabis alpina (Alpine rockcress).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabis.
OX NCBI_TaxID=50452;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AALP2;
RA Koch M., Mitchell-Olds T.;
RT "Evolutionary analysis of plastidic maturase K and nuclear chalcone
RT synthase and their utility for phylogenetic reconstructions within the
RT Brassicaceae.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS (BY SIMILARITY).
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
CC MITOCHONDRIAL INTRONS.
DR EMBL; AF144329; AAG43298.1; -.
DR InterPro; IPR000442; Intron_maturase2.
DR Pfam; PF01348; Intron_maturase2; 1.
DR Pfam; PF01844; MatK_Ar.1.
KW Chloroplast; mRNA processing.
SQ SEQUENCE 506 AA; 60472 MW; 3BD5743D758B9F23 CRC64;

Query Match 5.6%; Score 106.5; DB 8; Length 506;
Best Local Similarity 22.6%; Pred. No. 0.73;
Matches 29; Conservative 14; Mismatches 41; Indels 32; Gaps 1;

Qy 256 FWFVESGLTENHEGRKAYGAVGLISSPOELGHAFIDNVRLVPLLELDQIIRLPNTSTPOE 315
Db 2 YWFTVEYGLCKQNGEVKAYGAGLSSYGLVHLSLSDPETREFDPEAAAVQPYQDTYQP 61

Qy 316 TLF-----SIRHFDLVELTSKLEWLDQ 339
Db 62 VYFVSESFSDAKKPFYVAGIKRPFVFPDPTNSVEVLNPLKIQSGLEWVKDE 117

Query Match 5.6%; Score 106; DB 13; Length 129;
Best Local Similarity 25.0%; Pred. No. 0.16;
Matches 29; Conservative 14; Mismatches 41; Indels 32; Gaps 1;

Qy 256 FWFVESGLTENHEGRKAYGAVGLISSPOELGHAFIDNVRLVPLLELDQIIRLPNTSTPOE 315
Db 2 YWFTVEYGLCKQNGEVKAYGAGLSSYGLVHLSLSDPETREFDPEAAAVQPYQDTYQP 61

Qy 316 TLF-----SIRHFDLVELTSKLEWLDQ 339
Db 62 VYFVSESFSDAKKPFYVAGIKRPFVFPDPTNSVEVLNPLKIQSGLEWVKDE 117

RESULT 20
Q9GF55
```

```

Matches 68; Conservative 37; Mismatches 131; Indels 65; Gaps 12;

Qy 3 YCE-RTLDPKYILKIALKRLQSLISLFQNSQSLQRAYSTPYSYRIILOKENKEKQALAR 61
Db 190 YCNWKNFD-----IKKILINRPFLLYNSHVCE-----YESIFFLKRKSRSHURSTAY 239

Qy 62 HKCISILEPFKNL-----LFVH-----LLSLSKNQ-----REGCSTDMVVSTP----- 100
Db 240 EVLFEIRILFYAKIQHFLKVFVNPFALGLLKDPLFHYVRYHGKSLATKOTPLLMNKKW 299

Qy 101 FPNRLWYRLLSRFSMLKSYCPFFLDYLEAFGLSDFLDHOAVIKFFELTHFSYYPV 160
Db 300 FYFVNLWQYFVSQFQKIHINQLSKDNLEFLGYLSSRLNPLVVRSQLNSFLI--- 356

Qy 161 SGFVAPHQVLSLQDRYPIASVMRTLDKDNFSLTPDLIHDLGH-----VPLLHPSFSE 216
Db 357 -----DNIRIKLDNKIPISIIISLTGDKFC-----NLLGHPISKANW-TESSDS 401

Qy 217 FFINMGRFLTKVIEKVALPSKKQR-----IOTLQSNLIAIVRCFWFTVESGLI 265
Db 402 ILNRFVRIENISHYVSGSKKKHLYRIKYLRLCCVKTLLARKHKSTVRAFLKRLGSL 461

Qy 266 E 266
Db 462 E 462

RESULT 19
ID Q42428 PRELIMINARY; PRT; 129 AA.
AC Q42428;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Tyrosine hydroxylase (Fragment).
GN TH.
OS Lates calcarifer (Barramundi).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
OC Centropomidae; Lates.
OX NCBI_TaxID=8187;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BARRAMUNDI;
RA Collet C., Candy J., Sara V.;
RT "Tyrosine hydroxylase and insulin-like growth factor-II but not
RT insulin are adjacent in the teleost species barramundi (Lates
RT calcarifer).";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF007942; AAB64194.1; -.
DR HSSP; P04177; ITOH.
DR InterPro; IPR001273; Aaa_hydroxylase.
DR Pfam; PF00351; bioprotein_H; 1.
DR PRINTS; PR00372; FYHYDRLXASE.
DR NON TER 1
SQ SEQUENCE 129 AA; 14678 MW; 1AE29C7530AB4D78 CRC64;

Query Match 5.6%; Score 106; DB 13; Length 129;
Best Local Similarity 25.0%; Pred. No. 0.16;
Matches 29; Conservative 14; Mismatches 41; Indels 32; Gaps 1;

Qy 256 FWFVESGLTENHEGRKAYGAVGLISSPOELGHAFIDNVRLVPLLELDQIIRLPNTSTPOE 315
Db 2 YWFTVEYGLCKQNGEVKAYGAGLSSYGLVHLSLSDPETREFDPEAAAVQPYQDTYQP 61

Qy 316 TLF-----SIRHFDLVELTSKLEWLDQ 339
Db 62 VYFVSESFSDAKKPFYVAGIKRPFVFPDPTNSVEVLNPLKIQSGLEWVKDE 117
```

ID O9GF55 PRELIMINARY; PRT; 504 AA.  
 AC O9GF55;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Probable intron maturase (Maturase K).  
 GN MATK.  
 OS Cardamine amara (Large bitter-cress).  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta, eudicotyledons, core eudicots, Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Cardamine.  
 OK NCBI\_TaxID=50461;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Koch M., Mitchell-Olds T.;  
 RT "Evolutionary analysis of plastidic maturase K and nuclear chalcone  
 RT synthase and their utility for phylogenetic reconstructions within the  
 RT Brassicaceae.";  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II  
 CC INTRONS (BY SIMILARITY).  
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY  
 CC MITOCHONDRIAL INTRONS.  
 CC EMBL; AFI44337; AAG43306.1; -;  
 DR Interpro: IPR000442; Intron\_maturase2.  
 DR Interpro: IPR002866; MatK\_N.  
 DR Pfam: PF01348; Intron\_maturase2; 1.  
 DR Pfam: PF01824; MatK\_N; 1.  
 DR Chloroplast; mRNA processing.  
 SQ SEQUENCE 504 AA; 60332 MW; F557BF32927C7EF CRC64;

Query Match 5.6%; Score 106; DB 8; Length 504;  
 Best Local Similarity 22.8%; Pred. No. 0.81;  
 Matches 67; Conservative 36; Mismatches 115; Indels 76; Gaps 11;

OY 2 HYCERLDPKYLTKIALKQSLSPFONSQLOKAVSTPYSYRIILOKENEKQALAR 61  
 DB 214 HVCE-----YESIFFLKRSSHRSSTYEVLFERIFGYKIGHFLKVVY 258  
 OY 62 HKCISILEFPKNL--FVHLISLKNQ-----REGCSTDMVAVSTP-----FENNILVRLSSR 114  
 DB 250 -----IQHFVFNPNPAILGLKDPFIHYVHGRSILAKDPPLMNMKKKYFVNLM 304  
 OY 108 YRLSSRPSLWKSYPFRFLDYLEAFGLSDFLDHOAVIKFELETHFSYVSGFVA 167  
 DB 305 QCYFSWVFQSQKVKIKQLSKDNLEFLGYLSRLNPLVVRSCMLE-----NSFLIDN 356  
 OY 168 QYLSLIDPRYPPIASVARTLKDQNFSLTPDLIHDLLGH---VPMILHPSFSEFFIMNGR 223  
 DB 357 VRKLD--DSKIPISISIGSLAKKFC-----NVLGHSIRKVTW--THSSSDILNRFVR 406  
 OY 224 LFTKVIKVKQALPSKKOR-----IQTLQSNLIATVRCFWFTVESGLIE 266  
 DB 407 ICRNISHYSSGSKKFLYRIKYLRLCCVKTLARKHKSIVRAFLKRLGSLLE 460  
 RESULT 21  
 O9GF31 PRELIMINARY; PRT; 504 AA.  
 AC O9GF31;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Probable intron maturase (Maturase K).  
 GN MATK.  
 OS Cardamine penzancei.  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta, eudicotyledons, core eudicots, Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Cardamine.  
 OK NCBI\_TaxID=125588;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Koch M., Mitchell-Olds T.;  
 RT "Evolutionary analysis of plastidic maturase K and nuclear chalcone  
 RT synthase and their utility for phylogenetic reconstructions within the  
 RT Brassicaceae.";  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II  
 CC INTRONS (BY SIMILARITY).  
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY  
 CC MITOCHONDRIAL INTRONS.  
 CC EMBL; AFI44364; AAG43333.1; -;  
 DR Interpro: IPR000442; Intron\_maturase2.  
 DR Interpro: IPR002866; MatK\_N.  
 DR Pfam: PF01348; Intron\_maturase2; 1.  
 DR Pfam: PF01824; MatK\_N; 1.  
 DR Chloroplast; mRNA processing.  
 SQ SEQUENCE 504 AA; 60292 MW; 94C5F9B8282127A7 CRC64;

Query Match 5.6%; Score 106; DB 8; Length 504;  
 Best Local Similarity 22.3%; Pred. No. 0.81;  
 Matches 64; Conservative 41; Mismatches 120; Indels 62; Gaps 10;

OY 2 HYCERLDPKYLTKIALKQSLSPFONSQLOKAVSTPYSYRIILOKENEKQALAR 61  
 DB 214 HVCE-----YESIFFLKRSSHRSSTYEVLFERIFGYKIGHFLKVVY 258  
 OY 62 HKCISILEFPKNL--FVHLISLKNQREGCSTDMVAVSTP-----FENNILVRLSSR 114  
 DB 259 NNPFALLGLKD--PFIHY-----RHGRSILATKDPPLMNMKKKYFVNLMQCYFSW 311  
 OY 115 FSLMSKYPFRFLDYLEAFGLSDFLDHOAVIKFELETHFSYVSGFVAHQYLSLQ 174  
 DB 312 FQSQKVKIKQLSKDNLEFLGYLSRLNPLVVRSCMLE-----NSFLIDNVRKLD-- 361  
 OY 175 DRYPIASVARTLKDQNFSLTPDLIHDLLGH---VPMILHPSFSEFFIMNGRFLTKVIE 230  
 DB 362 DSKIPISISIGSLAKKFC-----NVLGHSIRKVTW--THSSSDILNRFVRCRNISH 413  
 OY 231 KVQALPSKKOR-----IQTLQSNLIATVRCFWFTVESGLIE 266  
 DB 414 YYSGSSKKKLYRIKYLRLCCVKTLARKHKSIVRAFLKRLGSLLE 460  
 RESULT 22  
 O9GF30 PRELIMINARY; PRT; 504 AA.  
 AC O9GF30;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Probable intron maturase (Maturase K).  
 GN MATK.  
 OS Cardamine rivularis.  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta, eudicotyledons, core eudicots, Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Cardamine.  
 OK NCBI\_TaxID=82338;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Koch M., Mitchell-Olds T.;  
 RT "Evolutionary analysis of plastidic maturase K and nuclear chalcone  
 RT synthase and their utility for phylogenetic reconstructions within the  
 RT Brassicaceae.";  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II  
 CC INTRONS (BY SIMILARITY).  
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY  
 CC MITOCHONDRIAL INTRONS.  
 CC EMBL; AFI44365; AAG43334.1; -;

DR InterPro; IPR000442; Intron\_maturase2.  
DR InterPro; IPR002866; MatK\_N  
DR Pfam; PF01348; Intron\_maturase2; 1.  
DR Pfam; PF01824; MatK\_N; 1.  
KW Chloroplast; mRNA processing.  
SQ SEQUENCE 504 AA; 60302 MW; 66AA07E9E2C48C64 CRC64;

Query Match 5.6%; Score 106; DB 8; Length 504;  
Best Local Similarity 22.3%; Pred. No. 0.81;  
Matches 64; Conservative 41; Mismatches 120; Indels 62; Gaps 10;

QY 2 HVCERTDPKYLKIALKURQBSLFFQNSOSLORAYSTPYVYRILOKENKEQALAR 61  
DB 214 HVC-----YESIFFLKRSLSRSTSYEVLPERIFFYKGIQHFLLKVFV 258  
QY 62 HVCISILEFFFNLLFVHLLSLKNOREGCSTDMAVVSTP-----FFNRLMYRLSSR 114  
DB 259 NNFPAILGLKLD-PFIHYV-----RYHGRSILATKDTPLLMNKKYKVFVNLWQCVFSW 311  
QY 115 FSLWKSYPFRFFLDYLEAFGLSDFLDHOAVIKFPELETHFSYVPVSGFVAPHOYLSLIQ 174  
DB 312 FQSKVHIKQLSKDNLEFLGLYSLRLNPLVRSQMLE-----NSFLIDNVRILK-- 361  
QY 175 DRYPIASVMRTLDKDNFSLTDLIDLGLH-----VPMLLHPSFSEFFINMGLRFTKVIE 230  
DB 362 DSKTPIGSIIGSLAKKFC-----NVLGHPSKVIV-THSDSDILNRFVRCINISH 413  
QY 231 KVQALPSKKOR-----IQTLQSNLIAIVRCFWFTVESGLIE 266  
DB 414 YYSGSKKKLYRIKYLRLCCVKTLARKHKSTVRAFLLKRLGSGLLIE 460

RESULT 23  
Q95F33 PRELIMINARY; PRT; 513 AA.  
AC Q95F33;  
DT 01-DEC-2001 (T-EMBLrel. 19, Created)  
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)  
DE Probable intron maturase (Maturase K).  
GN MATK.  
OS Muhlenbergia sylvatica.  
OC Chloroplast  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
OC Chloridoideae; Eragrostidae; Muhlenbergia.  
OX NCBI\_TaxID=160557;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hilu K.W., Alice L.A.;  
RT "A Phylogeny of Chloridoideae (Poaceae) Based on matK Sequences.";  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II  
CC INTRONS (BY SIMILARITY).  
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,  
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY  
CC MITOCHONDRIAL INTRONS.  
DR EMBL; AF312355; AAK60046.1; -;  
DR InterPro; IPR000442; Intron\_maturase2.  
DR InterPro; IPR002866; MatK\_N  
DR Pfam; PF01348; Intron\_maturase2; 1.  
DR Pfam; PF01824; MatK\_N; 1.  
KW Chloroplast; mRNA processing.  
SQ SEQUENCE 513 AA; 61468 MW; 8D6AAD01B14BDE8 CRC64;

Query Match 5.5%; Score 104; DB 8; Length 513;  
Best Local Similarity 18.6%; Pred. No. 1.2;  
Matches 67; Conservative 67; Mismatches 124; Indels 102; Gaps 17;

QY 22 QSLSLFFQNSQSLQOR-----AYSTPYSYRRILOKENKEQALARKHKCISILEFFKXLLFV 77  
DB 199 KSIFLFSKDNKRLSRFLYNYFVSEYEFFLLRLKQS-----SCLRLTSYGTFLERI 249

QY 78 HLLSLKNOREGCSTDMAVVSTPFFNRLMY-----RLLSRFSFL-----W 118  
DB 250 HFFRKMEH-----FGVMYEGFPQKTIWFFMEPLMHVYRQKVLASKGTLLFQKKW 301  
QY 119 KSYCPFRFFLDYLEAFGLSDFLDHOAVIKFPELETHFSYVPVSGFVAPHOYLSDLDQRYF 178  
DB 302 KSYLVNFSQYFFSFWAQPORICLNQLTNSCDFLGYRSNVPIINTFLVITNQL-----ENFF 357  
QY 179 PIASVMRTLDKDNFSLTDLIDLGLHVPMLLHPSFSEFFINMGLRFTK-----V 228  
DB 358 LIATRMKKFD-TTAPATP-----LIGSL-----SKAQFCTGSHGISPIWTDLSDWDI 405  
QY 229 IEKVOAL-----PSKORIOTLOSNIATVRCFWFTVESGLIENHEGRKAYGAVL 278  
DB 406 LDRFGTRICNRLPHYHSGSKQTLRLK-----YILR-----LSCARTLARKHK-----STV 452  
QY 279 ISSQEQLGHAFIDNVRVLPLELDQIIRLPNTSTPQETLSIRHFDDELVELTSKLEWMLD 338  
DB 453 RTFMQLRGSVLEBEFT-----EBEQVSLMFSKTT-----HFSFHGSYSERI-WYLD 499

RESULT 24  
Q95F29 PRELIMINARY; PRT; 513 AA.  
AC Q95F29;  
DT 01-DEC-2001 (T-EMBLrel. 19, Created)  
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)  
DE Probable intron maturase (Maturase K) (fragment).  
GN MATK.  
OS Enneapogon glaber.  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
OC Chloridoideae; Pappophoreae; Enneapogon.  
OX NCBI\_TaxID=160589;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hilu K.W., Alice L.A.;  
RT "A Phylogeny of Chloridoideae (Poaceae) Based on matK Sequences.";  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II  
CC INTRONS (BY SIMILARITY).  
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,  
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY  
CC MITOCHONDRIAL INTRONS.  
DR EMBL; AF312360; AAK60051.1; -;  
DR InterPro; IPR000442; Intron\_maturase2.  
DR InterPro; IPR002866; MatK\_N  
DR Pfam; PF01348; Intron\_maturase2; 1.  
DR Pfam; PF01824; MatK\_N; 1.  
KW Chloroplast; mRNA processing.  
FT NON\_TER 513  
SQ SEQUENCE 513 AA; 61341 MW; CD93D8582ECC3682 CRC64;

Query Match 5.5%; Score 104; DB 8; Length 513;  
Best Local Similarity 20.3%; Pred. No. 1.2;  
Matches 72; Conservative 60; Mismatches 131; Indels 92; Gaps 16;

QY 22 QSLSLFFQNSQSLQOR-----AYSTPYSYRRILOKENKEQALARKHKCISILEFFKXLL 74  
DB 199 KSIFLFSKDNKRLSRFLYNYFVSEYEFFLLRLKQS-----SCLRLTSYGTFLERI 249  
QY 75 LFWHLLSLKNOREGCSTDMAVVSTPFFNRLMY-----RLLSRFSFL----- 117  
DB 250 IFSRME-----HFGVMYEGFPQKTIWFFMEPLMHVYRQKVLASKGTLLLK 298  
QY 118 ---WKSYPFRFFLDYLEAFGLSDFLDHOAVIKFPELETHFSYVPVSGFVAPHOYLS---L 172  
DB 299 KWKXSYLVNFSQYFFSFWAQPORICLNQLTNSCDFLGYRSNVPMQFLVRNQMLNSFL 358  
QY 173 LQ-----DRYPTIASVMRTLDKDNFSLTDLIDLGLH-VPMLLHPSFSEFTI--NMGR 223

Db 359 ITRAKKPTTVPVTSLSGSKAKFCG-----LGHISKPIWTDLSMDILDRGR 411  
Qy 224 LFTKVEKVOALPSKKRIQTLQSNLIAIVRCFMFVESGLIENEGKAYAVLISSPQ 283  
Db 412 ICRNLPH-YHSGSKKQLYRLK---YLIR---LSCARTLARKH-----STVTFMQ 457  
Qy 284 ELGHAFTDNVRLPLELDOIIRLPNTSTPOETLFSIRHPEDELVELTSKLEWMLD 338  
Db 458 RLGSVLEBEFT---EEQVPSLMFTKTT-----HF-SPHGHSERIWYLD 499

## RESULT 25

Q95F39 PRELIMINARY; PRT; 513 AA.  
ID 095F39  
AC 095F39; 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
DE Probable intron maturase (Maturase K).  
GN MATK.  
OS Monothochloe littoralis.  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
OC Chloridaceae; Eragrostidae; Monothochloe.  
ON NCBI\_TaxID=160556;  
RX  
RP SEQUENCE FROM N.A.  
RA Hilu K.W., Alice L.A.;  
RT "A Phylogeny of Chloridoideae (Poaceae) Based on matK Sequences";  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
CC -I- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II  
CC INTRONS (BY SIMILARITY).  
CC -I- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,  
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY  
CC MITOCHONDRIAL INTRONS.  
DR EMBL: AF312349; AAK60040.1; -;  
DR InterPro: IPR000442; Intron\_maturase2.  
DR Pfam: PF01348; Intron\_maturase2; 1.  
DR Pfam: PF01824; MatK\_N; 1.  
KM Chloroplast; mRNA processing.  
SQ SEQUENCE 513 AA; 61512 MW; FA97488C266572D8 CRC64;

Query Match 5.5%; Score 103.5; DB 8; Length 513;  
Best Local Similarity 21.0%; Pred. No. 1.3; Indels 89; Gaps 19;  
Matches 75; Conservative 59; Mismatches 134;

Qy 12 YLTKIALKRLQSLIFPONSQSLQRAYSTPYSYVRIILQENKEKQALARKHCISILEFF 71  
Db 202 FLK---KENRRLSRFLYNS-----YVSEYEFLLFLRKOS-----SCLRLA 240  
Qy 72 KLLFVHLLSLKNQRECSIDMAVSTPFENLWY-----RLSSRSL 117  
Db 241 SSGTFLERIHFSRKMEH-----FCWYRGFFRTIWFMDPLMHVYRQGYKILASKCTL 295  
Qy 118 -----WKSYPRLFVLEAFGLSDFLDHQAVIKFPELETHSYVPYSGFAAPROYSL 172  
Db 296 LFOKKMKYLVNFSQYFSSPTQQRIRANQLNSCFDVLGRSSVPIINTFLVKNQML-- 353  
Qy 173 LQDRYFPIASVMTLDKDNBSLTPDLIHD-----LGH-VPMILHSESEPT--NM 221  
Db 354 --ESFFPIATMKKPD--TPAPATP-LIGSSSKQFCTGLGHPISKPIWTDLSMDILDRF 409  
Qy 222 GLFTKVEKVOALPSKKRIQTLQSNLIAIVRCFMFVESGLIENEGKAYAVLISS 281  
Db 410 GICRNLPH-YHSGSKKQLYRLK---YLIR---LSCARTLARKH-----STVTF 455  
Qy 282 POLGHAFTDNVRLPLELDOIIRLPNTSTPOETLFSIRHPEDELVELTSKLEWMLD 338  
Db 456 MRLGSVLEBEFT---EEQVPSLMFTKTT-----HFSFGHSERI-WYLD 499

## RESULT 26

Q9VP19 PRELIMINARY; PRT; 1782 AA.  
ID 09VP19  
AC 09VP19; 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)  
DE CG11376 protein.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Ephydroidea; Drosophilidae; Drosophila.  
ON NCBI\_TaxID=7227;  
RX  
RP SEQUENCE FROM N.A.  
RC SPMRAIN=BRKRELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoekins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abell J.F., Ashbyant A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
RA Burlie K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Ewing-Delmon S., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fodde C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Foster A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heilmann T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Idegian C.,  
RA Jalaal M., Kalush F., Karpen G.H., Ke, Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Lin X., Matrei B., McInosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Mostrel A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Murthy D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Maasman D.A., Weinstein G.M., Weisenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RT "The genome sequence of Drosophila melanogaster";  
RL Science 287:2185-2195 (2000).  
DR EMBL: A8003590; AAF51561.1; -;  
DR FlyBase: FBgn003216; CG11376.  
SQ SEQUENCE 1782 AA; 202676 MW; 3C80F5EC17D5E1B CRC64;

Query Match 5.5%; Score 103.5; DB 5; Length 1782;

Best Local Similarity 25.0%; Pred. No. 6.1; Indels 96; Gaps 15;  
Matches 73; Conservative 32; Mismatches 96;

Qy 67 ILRF-----FKNLLFVHLLSLKNQRECSIDMAV-----VSTPFENRLWYR 109  
Db 574 ILRFPOSALYNPHYSYRNLLFVSPKELNFSRASANNIAVRLQWAGSTPKDAVAYI- 632  
Qy 110 LLSRSFLWKSYPRLFVLEAFGLS-----DFPDHQAIVKFELETH-----FSYYP 159  
Db 633 -----GKSSCPKF--STEAFTAVVYHNKCSFSDERIKIALPASIKQHHLHFTIYH 681  
Qy 160 VSGFVAP-----HGYLSLDQRYRPIAS-----VMTLDKDNBSLTPDLIHDLL 203

Db VSCQKQDQLQPSVETPIGYTWLPLEDKLKFGEFNLPMVSPENYSFIPPNVH--L 739  
Qy 204 CHVFWLLHPSFSEFFINMGRLETKVIEKQVALPSKKORIOTLOSNIIA-IVRCFWFTVES 262  
Db 740 PGIKWL-----DNHRAVSINVEAWTA-----IHTLDSFLDRFFELICEYLDTRN 783  
Qy 263 GLIENHGRKAYGAVLISSPQELGHAFID-----NVRVLPLELDQIIRL 306  
Db 784 --IPSHIGEN-----NIETELKCLLDIEYANREPLVRHLPIVLDKLIEL 826

## RESULT 27

Q9GF41 ID Q9GF41 PRELIMINARY; PRT; 504 AA.  
AC Q9GF41, 2001 (TREMREL. 16, Created)  
DT 01-MAR-2001 (TREMREL. 16, Last sequence update)  
DT 01-MAR-2002 (TREMREL. 20, Last annotation update)  
DE Probable intron maturase (Maturase K).  
GN MATK.  
OS Aubrieta deltoidea.  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Aubrieta.  
OX NCBI\_TaxID=81984;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Koch M., Mitchell-Olds T.;  
RT "Evolutionary analysis of plastidic maturase K and nuclear chalcone  
RT synthase and their utility for phylogenetic reconstructions within the  
RT Brassicaceae.";  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II  
CC INTRONS (BY SIMILARITY).  
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,  
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY  
CC MITOCHONDRIAL INTRONS.  
DR EMBL; AF144352; A043321.1; -.  
DR InterPro; IPR000442; Intron\_maturase2.  
DR InterPro; IPR002866; MatK\_N.  
DR Pfam; PF01348; Intron\_maturase2; 1.  
DR Pfam; PF01824; MatK\_N; 1.  
KW Chloroplast; mRNA processing.  
SQ SEQUENCE 504 AA; 60023 MW; 4620F813810B3AD8 CRC64;

Query Match 5.5%; Score 103; DB 8; Length 504;  
Best Local Similarity 21.9%; Pred. No. 1.5;  
Matches 66; Conservative 37; Mismatches 113; Indels 86; Gaps 12;  
Qy 8 LDPKYLKIALKRLQSLSPFONSQSLORAYSTPYVYRILOKENKEKQALARKHCISI 67  
Db 202 LNPRFL-----LFLYNHVC-----YESIFFLR-----KRSSPLRSTAYEV 239  
Qy 68 LEFFKNLIV-----HLLSLKQ-----REGCTDMAVVSTP-----100  
Db 240 L--FERILFYGIHFLKLVFNVNFLTIPGLLKDPFLVYRHGKSILATKDTPLMKNWK 297  
Qy 101 FPNRLWYLLSSRFSLWKSICYCPRFDFLYLEAFGLLSDFLDQHVAKVFFLETHFSYYPV 160  
Db 298 FFFVNLWCQYFSVWFQSKVNIQLSKNLEFLGYLSRLNPLVRSQMLENSFLI---354  
Qy 161 SGFVAPHQYLSLQDRYFFPIASVMRTLDKDNFSLTPDLIHLGLH-----VPWLLHPSFE 216  
Db 355 -----DNIRIKLDSKIPISSTMGSLANDKFC-----NVLGHPISKATWTDSDDFD- 399  
Qy 217 PFINMGRFLTKEVIEKQVALPSKKOR-----IQTLOSNIIAIVRCFWFTVESGL 264  
Db 400 -ILNRFVRCIRNHHYSGSKKKNLYRIKTYILRCCVCKTLARKHKSTVRAFLKRVGSL 458  
Qy 265 IE 266

Db 459 VE 460  
RESULT 28  
Q9TIB8 ID Q9TIB8 PRELIMINARY; PRT; 513 AA.  
AC Q9TIB8;  
DT 01-MAY-2000 (TREMREL. 13, Created)  
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)  
DT 01-MAR-2002 (TREMREL. 20, Last annotation update)  
DE Probable intron maturase (Maturase K).  
GN MATK.  
OS Bouteloua curtipendula.  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
OC Chloridoideae; Cynodonteae; Bouteloua.  
OX NCBI\_TaxID=110875;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hilu K.W., Alice L.A.;  
RT "Phylogenetic relationships in subfamily Chloridoideae (Poaceae) based  
RT on matk sequences: A preliminary assessment.";  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II  
CC INTRONS (BY SIMILARITY).  
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,  
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY  
CC MITOCHONDRIAL INTRONS.  
DR EMBL; AF144578; AAF20334.1; -.  
DR InterPro; IPR000442; Intron\_maturase2.  
DR InterPro; IPR002866; MatK\_N.  
DR Pfam; PF01348; Intron\_maturase2; 1.  
DR Pfam; PF01824; MatK\_N; 1.  
KW Chloroplast; mRNA processing.  
SQ SEQUENCE 513 AA; 61519 MW; 223EEA9CEA96DFEB CRC64;  
Query Match 5.5%; Score 103; DB 8; Length 513;  
Best Local Similarity 22.0%; Pred. No. 1.5;  
Matches 78; Conservative 59; Mismatches 127; Indels 90; Gaps 22;  
Qy 22 QSLSLFONQSQSLOR-----AYSTPYVYRILOKENKEKQALARKHCISILEFFKNLLFV 77  
Db 199 KSIFLFSKENKRLSRFLYNSYVSEYEFLLFLRKQS-----SCLRTSSGTGR 246  
Qy 78 HLLSLKNQREGCGTDMAVVSTPFFNRNLW-----YR---LLSRFSL-----W 118  
Db 247 ERIFFSRKMGEH-----FGVMYPGFELKTLIFMDPLMHVRYQGVILASKGTLFLFQKW 301  
Qy 119 KSYCPREFFLDYLEAFGLLSDFLD-HQAVIKPFELETHFSYYPVSGFVAPHQYLSLQDRY 177  
Db 302 KSYLVN-FSQYFFSFWIQPQIRINQLTNSCFDLGVHSSVPINTFLVRNQL-----ENF 356  
Qy 178 FPIASVMRTLDKDNFSLTPDLIHLG-----LGH-VPWLLHPSSEFEFI--NMGRFLT 226  
Db 357 FLIATRMKKFD-TTVATP-LIGSLSKAQFCCTGLGHPISKPIWTDLSDWILDREFGICR 414  
Qy 227 KVIEKQVALPSKKORIOTLOSNIIAIVRCFWFTVESGLIENHGRKAYGAVLISSPOEIG 286  
Db 415 NLFH-YHSGSKKQTYLRUK-----YILR---LSCARTLARKHK-----STVTFMQRLG 460  
Qy 287 HAFIDNVRVLPLELDQIIRLPFNVTSTPQSTLFSIR--HFDDELVELTSKLEWMLD 338  
Db 461 SVFLEEFFT---EEQVFSLMF-----PRATHFSFHGSHSERI-----WYLD 499  
RESULT 29  
Q95F60 ID Q95F60 PRELIMINARY; PRT; 513 AA.  
AC Q95F60;  
DT 01-DEC-2001 (TREMREL. 19, Created)  
DT 01-DEC-2001 (TREMREL. 19, Last sequence update)  
DT 01-MAR-2002 (TREMREL. 20, Last annotation update)



```
DR Pfam; PF01348; Intron_maturas2; 1.
DR Pfam; PF01824; MatK_N; 1.
KW Chloroplast; mRNA processing.
SQ SEQUENCE 513 AA; 61546 MW; 08330B6C2282C755 CRC64;

Query Match
Best Local Similarity 20.5%; Pred. No. 1.5;
Matches 72; Conservative 62; Mismatches 133; Indels 84; Gaps 18;

QY 22 QSLSLFFQNSQSLR-----AYSPPSYVYRIILQENKEKQALARKHCISLIEFFKNLFPV 77
DB 199 KSIELEKKEKRLSRFLNYSYSEYEFLLFLRKQS-----SCLRLASSGTF 246
QY 78 HLLSLSKNQREGCSTDMVAVSTPFNNLWY-----RLLSRSL-----W 118
DB 247 ERIHSRKMEH-----FGWYPCFFRKIWFMDPLMHVYVYQGVILASKGTLFQKXW 301
QY 119 KSYCPRFLDYLEAFGLSDFLDHQAVIKFELETHFSYVPVSGFVAPHQYLSLLODRYF 178
DB 302 KSYLVNLSQFFSWFQTPQRIQLNQTNSCDFLGIRSVFINTFLVRNQML-----ESFF 357
QY 179 PIASVMTLKDNLFTPLDIHDL-----LGH-VPLMLHPSFSFEFFI--NMGRLFTK 227
DB 358 PIATRMKKFD--TTATATP-LIGSLSKAQCCTGLGHPISKPIWTDLSDWDILDRFGICRN 415
QY 228 VIEKQVLPSPKRIQIOTLQSNLAIIVRCFWFTVESGLTENHGRKAYGAVLISSPQELGH 287
DB 416 LFH-YHSGSSKKRTYLRK-----YILR---LSCARTLARKHK-----STVTFMQRLGS 461
QY 288 AFIDNVRLPLELDQIIRLPFNSTPQETLFSIRHDELVELTSKLEWMLD 338
DB 462 VLEBEFFT---EEQVFLMFAKTT-----HFSPHGSHSRI-WYLD 499

RESULT 32
Q9VM23
ID Q9VM23 PRELIMINARY; PRT; 1154 AA.
AC Q9VM23
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE NimaC protein.
GN NINAC OR CG5125.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton R.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu U., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman M.R., Bouck J., Brockstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Caudle E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo J.M., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
```

DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Probable intron maturase (Maturase K).  
 GN MATK.  
 OS Arabis alpina (Alpine rockcress).  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabis.  
 ON NCBI\_TaxID=50452;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ALP1.  
 RA Koch M., Mitchell-Olds T.;  
 RT "Evolutionary analysis of plastidic maturase K and nuclear chalcone  
 RT synthase and their utility for phylogenetic reconstructions within the  
 RT Brassicaceae.";  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II  
 CC INTRONS (BY SIMILARITY).  
 CC -1- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS, BY  
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY  
 CC MITOCHONDRIAL INTRONS.  
 DR EMBL; AF144328; AAG43297.1; -  
 DR InterPro; IPR000442; Intron\_maturase2.  
 DR Pfam; PF01348; Intron\_maturase2; 1.  
 DR Pfam; PF01824; MatK\_N; 1.  
 DR Chloroplast; mRNA processing.  
 KW SEQUENCE 506 AA; 60662 MW; 275A64406CEB9525 CRC64;  
 SQ

Query Match 5.4%; Score 102; DB 8; Length 506;  
 Best Local Similarity 22.9%; Pred. No. 1.8;  
 Matches 67; Conservative 35; Mismatches 117; Indels 74; Gaps 11;

QY 2 HYCERTLPKYLKALRLROSLFQNSQSLQRAYSTPYSYVYIIQKKEKQALR 61  
 DB 216 HYCE-----YESIFFLARKSSHLRSTAYEVLPERILF-----YAKIQ 253  
 QY 62 HNCISLEFFKMLFVHLISLKNQ-----REGGCTDMAVSTP-----FERNLMYR 109  
 DB 254 H-----FLKVFVN-NFPAIIGLLKDPFLHYVYHGXSLTKDPLANKKFFVYVLMQF 308  
 QY 110 LLSRFLSKSYCPRRFLDYLEAFGLISDFLDHQAIVKFELETHPSYVSGVAPHOY 169  
 DB 309 YFSVWFQSKTHINQSLKDNLEFLGYLSLRNPLVVRQOMLENSFLI-----DN 358  
 QY 170 LSLQDRYFPPIASVWRLDKNFSLTPLDILHDLGH---VPMLLHPSFSEFFINMGRLF 225  
 DB 359 IRIKLDNKIPISISIIIGSLTKDKFC-----NLGHPISKAWMT--ESSDSIDINRFLRI 409  
 QY 226 TVYIEKVALPSKKQR-----IOTLOSNIATVRCFMTVESGLIE 266  
 DB 410 CRNISHYSSGSSKKHLYRIKYLRLCCVKTLLARKKSTVRAFLKRLSGGLE 462  
 RESULT 34  
 ID 09MV23 PRELIMINARY; PRT; 519 AA.  
 AC 09MV23;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Probable intron maturase (Maturase K) (Fragment).  
 GN MATK.  
 OS Oryza meyeriana.  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Euphorbiaceae; Oryzaceae; Oryza.  
 ON NCBI\_TaxID=83307;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=20056256; PubMed=10588717;  
 RA Ge S., Sang T., Lu B.R., Hong D.Y.;  
 RT "Phylogeny of rice genomes with emphasis on origins of allotetraploid  
 RT species.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:14400-14405(1999).  
 CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II  
 CC INTRONS (BY SIMILARITY).  
 CC -1- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,  
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY  
 CC MITOCHONDRIAL INTRONS.  
 DR EMBL; AF148673; AAF37181.1; -  
 DR InterPro; IPR000442; Intron\_maturase2.  
 DR Pfam; PF01348; Intron\_maturase2; 1.  
 DR Pfam; PF01824; MatK\_N; 1.  
 DR Chloroplast; mRNA processing.  
 KW NON\_TER  
 FT NON\_TER 519 519 1  
 SQ SEQUENCE 519 AA; 62257 MW; 0148830573EB14C1 CRC64;  
 Query Match 5.4%; Score 101.5; DB 8; Length 519;  
 Best Local Similarity 24.5%; Pred. No. 2;  
 Matches 96; Conservative 55; Mismatches 148; Indels 93; Gaps 26;

QY 1 VHYCERTLPKYLKALRLKLRQ-----SLSL--PFQNSQSLQRAYSTPYSYVYIIQ 50  
 DB 165 LHYLSHIEIYPIHDLILLQIQRIDQVPSHLRPFLLNYVSSMNSFITSMKSI-LILK 223  
 QY 51 KENKQALARHKCISILEFFKMLFVH---LISLKN-----REGGCTDM--AVYST 99  
 DB 224 KENKRLPRFLNYSYSEYEFF--LFLKQSSGLRLSSGTPLERIHFBRKKEHGVWVP 281  
 QY 100 PFERNLMY-----RLSSRFL-----KSYCPRRFLDYLEAFGLISDPL 140  
 DB 282 AFRKTVYFPMQDMHYVYQYQKAILAKSGTLKKKCYLYRLM-QYSPSFQOSRI 340  
 QY 141 D-HQAVIKFELETHSYVSGVAPHOYISLQDYPPIASVWRTDKNFSLTPLDI 199  
 DB 341 HLMQANSCBDFLGTSVPIVNSLVNRQM---ENSLFDTRMKFD--TKVPATP-LI 394  
 QY 200 HDLI-----GH-VPMLLHPSFSEFT--NMGRFTKYIEKVALPSKRIOTLOS 248  
 DB 395 GSLAKAQFTGSGHPISKPIWTDSLWDIDLRFERICRNLFH-YHSGSSKKKTLVRLK-- 451  
 QY 249 LIAIVRCFMTVESGLIENHEGRKAYGAVLISFQELGNAFIDNVVRLPLELDQIIRLPF 308  
 DB 452 --YILR--LSCARTLARKHK-----STVRAFPQWIGSVLEBEFT--EEEOVFSIMF 497  
 QY 309 NTSTPQETLFSIR--HFDLVELTSKLEWMD 338  
 DB 498 ----AKTYFFSFRGSHSDRI-----WYLD 517  
 RESULT 35  
 ID 08T2D4 PRELIMINARY; PRT; 1615 AA.  
 AC 08T2D4;  
 DT 01-JUN-2002 (Tremblrel. 21, Created)  
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
 DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Hypoetical 192.5 kDa protein.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostelidae; Dictyostelium.  
 ON NCBI\_TaxID=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX4.  
 RA Gloeckner G., Eichinger L., Szafrański K., Pachebat J., Dear P.,  
 RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,  
 RA Tunggal B., Cox B., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;  
 RT "Sequence and Analysis of Chromosome 2 of Dictyostelium.";  
 RT Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.





RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003462; AAF47145.1; -;  
 DR Flybase: FBgn0034964; CG3173.  
 SQ SEQUENCE 2042 AA; 233746 MW; 9144DB7B51F52BES CRC64;

Query Match 5.3%; Score 101; DB 5; Length 2042;

Best Local Similarity 21.1%; Pred. No. 12;

Matches 78; Conservative 49; Mismatches 122; Indels 120; Gaps 19;

QY 48 ILKENKEKQALAHKCSILEPKNL-----LFVHLISLKNREGC 90  
 DB 879 IINENSADAL-----VNFRLNIFDGNLSHQVCELDYFRLSLSTVKQSVAA 930  
 QY 91 STDAAVY---STPFNNLWYRLSS-----RFSLMKSCYCPFLDY 129  
 DB 931 LSGKTIIRHSGDF--ENEM--LKSLOQIPHYEVKPIIIPQLAAACOVNCELLIMAY 986  
 QY 130 LE--AFGL--LSDFLDH-----QAVIKFELETH-----FQYVSG-----F 163  
 DB 987 IQFTATLNDPVENMLDHYIDMAQGLIVERSTFQGIISQSDYIVFDENRITQTLKLP 1046  
 QY 164 VAPHOYSLLODRYFPPIASVWRITLDKDNFSLPDLI--HDLLGHVPMWLHPSESEFFIN 220  
 DB 1047 VMFNYYIILKREYHEPY-----EMTEYFDLIMVQFDGVLPL--LH-----IN 1087  
 QY 221 MGRLETFVIEKVALPSKRIOTLOSMLIAIVRCFWFTVESGLIENHEGRKAYGAVILIS 280  
 DB 1088 ITHAF--IILTYSNSNMPSISPLD-----YMPF-----PGRPAPAVAFPS 1127  
 QY 281 SPQELGHAFFID--NVRVLPLELDQIIRLPNTSTPOETLFSIRHDELVEVTSKLEWMLD 338  
 DB 1128 MPQEOVQLPMLKTKMIRSSVDRLEALNLDLPQIVLFVQNGFTVYNSMSKILMLD 1187  
 QY 339 OGLLESIPL 347  
 DB 1188 TAVLEQFDL 1196

RESULT 38

P97517 PRELIMINARY; PRT; 102 AA.

ID P97517 P70468;  
 AC P97517, P70468;  
 DT 01-MAY-1997 (Tremblrel. 03, Created)  
 DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)  
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
 DE Tyrosine hydroxylase (fragment).  
 OS *Phodopus sungorus* (Striped hairy-footed hamster) (Djungarian hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC *Phodopus*.  
 OC NCBI\_TaxID=10044;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=HYPOPHALAMUS;  
 RA Bockman J.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: Y09294; CAA70476.1; -;  
 DR HSSP: P04177; ITOH;  
 DR InterPro: IPR001273; Aaa\_hydroxylase.  
 DR Pfam: PF00351; bioperlin\_H.1.  
 DR PRINTS: PR00372; FYMHYDKLASE.  
 FT NON\_TER 1  
 FT NON\_TER 102  
 SQ SEQUENCE 102 AA; 11433 MW; 52241A9D5DB27795 CRC64;

Query Match 5.3%; Score 100.5; DB 11; Length 102;  
 Best Local Similarity 29.3%; Pred. No. 0.35; 43; Indels 7; Gaps 1;  
 Matches 27; Conservative 15; Mismatches 43; Indels 7; Gaps 1;

QY 234 ALPSKRIQTLQSNLIAIVRCFWFTVESGLIENHEGRKAYGAVILISPOELGHAFFIDNV 293  
 DB 10 SLGASDEELKST-----VMPFVFGKQNGSLKAYGAGLSTYGLHLSSEEP 62  
 QY 294 RVLPLELDQIIRLPNTSTPOETLFSIRHDE 325  
 DB 63 EVRAFPEAAMQPYODQTYQPVYVSSFSF 94

RESULT 39

Q9D0D1 PRELIMINARY; PRT; 2212 AA.

ID Q9D0D1  
 AC Q9D0D1;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE L protein.  
 GN L.  
 OS Ebola virus (strain Zaire Mayinga) (Ebo).  
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales; Filoviridae;  
 OC Ebola-like viruses.  
 OC NCBI\_TaxID=128952;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAYINGA;  
 RC MEDLINE=9911703; PubMed=10073695;  
 RA Velichov V.E., Velichova V.A., Chepurnov A.A., Blinov V.M., Dolnik O.,  
 RA Netesov S.V., Feldmann H.;  
 RT "Characterization of the L gene and 5' trailer region of Ebola  
 RT virus.";  
 RL J. Gen. Virol. 80:355-362(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAYINGA;  
 RC PubMed=11062045;  
 RA Velichov V.E., Chepurnov A.A., Velichova V.A., Ternovoj V.A.,  
 RA Kleen H.D.;  
 RT "Molecular Characterization of Guinea Pig-Adapted Variants of Ebola  
 RT Virus.";  
 RL Virology 277:147-155(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAYINGA;  
 RA Velichov V.E.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: A8272001; X824017.1; -;  
 DR InterPro: IPR002877; Ftsd;  
 DR InterPro: IPR001016; Viral\_RNA\_pol\_L.  
 DR Pfam: PF01728; Ftsd; 1.  
 DR Pfam: PF00946; Paramyx\_RNA\_pol; 1.  
 SQ SEQUENCE 2212 AA; 252692 MW; 5834C0A291603DAB CRC64;

Query Match 5.3%; Score 100.5; DB 12; Length 2212;

Best Local Similarity 23.1%; Pred. No. 14;

Matches 62; Conservative 42; Mismatches 95; Indels 69; Gaps 13;

QY 13 ILKIALKLRQSLPFPNSQSLQRAVSTPYVYRIILQKENKEKQALAHKCSILEPKNL 72  
 DB 355 LIRLEMPQQLCELF-----SIQKMGHPVLHSETAIQVKKATVLLKALRPVIFETV- 408  
 QY 73 NLFPVHLISLKNREGCSTDAAVVS-----TP-----FFNNLWYRLSSRFSLMKSCYCP 124  
 DB 409 ---CVFKYSIAKIVPDSQGSWSVTSDBNLPLGNSYIKKNQPPPLMTELMEFY--- 462  
 QY 125 FPLDYLEAPGL-----LSDFLDHQAVI-----KPELETHNPSYVVGSPVAPHOYSL 172  
 DB 463 -HDDHPLPFTKTIISLIFIKDKATVAVERTCMADVBN-----VLGINPPIKFSYK 514

QY 173 LODRYFFPIASVMRTLDKDNFSLTPDLIHLGHVPL--LHPSFSEFF-----INMR 223  
DB 515 RVPEQF-----LEQNF-----IENVLSYAQKLYLLPOYRNFSLKEKELNVR 561  
QY 224 LFTKVIKQVALPSKKORIQTLOSLNIA 251  
DB 562 TFGK-----LPYTRNVQTLCEALLA 582

RESULT 40  
Q9TI97 PRELIMINARY; PRT; 513 AA.  
AC Q9TI97;  
DT 01-MAY-2000 (T-EMBLrel. 13, Created)  
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)  
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)  
DE Probable intron maturase (Maturase K).  
GN MATK  
OS Sporobolus indicus (smut grass).  
OC Chloroplast  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
OC Chloridoideae; Eragrostidae; Sporobolus.  
OX NCBI\_TaxID=38731;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hilu K.W., Alice L.A.;  
RT "Phylogenetic relationships in subfamily Chloridoideae (Poaceae) based  
on matk sequences: A preliminary assessment."  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II  
INTRONS (BY SIMILARITY).  
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,  
AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY  
MITOCHONDRIAL INTRONS.  
CC EMBL: AF144601; AAF20357.1; -;  
DR InterPro: IPR000442; Intron\_maturase2.  
DR InterPro: IPR002866; MatK\_N  
DR Pfam: PF01348; Intron\_maturase2; 1.  
DR Pfam: PF01824; MatK\_N; 1.  
KW Chloroplast; mRNA processing.  
SQ SEQUENCE 513 AA; 61561 MW; BF6BB16048B3ECC6A CRC64;

Query Match 5.3%; Score 99.5; DB 8; Length 513;  
Best Local Similarity 23.3%; Pred. No. 3;  
Matches 91; Conservative 54; Mismatches 153; Indels 93; Gaps 20;  
QY 5 ERTLPKYLKALK-----LRQSLFFQNSQLO--RAYSTPYSYR-----46  
DB 142 EKFLHLVLSHIEIPIPHPEILLVQLLEYRIQDVPSLHLLRFFLLNYYSNNLSITSMKSI 201  
QY 47 ILOKENKEQALARKHCISILEFFKNLLFVHLLSLKN-----QREGCSTDN--A 95  
DB 202 FLLKKENKRLPRFLYNSVSEYEFF--LLFLRKQSSSLRLISSGTFLELRIHFSMKMEHFG 259  
QY 96 VYSTPFFNRNLWY-----RLSSRFSL-----WKSVCPRFFLDYLAFLGL 136  
DB 260 VMYRPFQKTLWPFMDPLMHVRYQKAILASKGTLLKLLKWKSYLVNFSQYFLSPWTP 319  
QY 137 SDFLDHQAVIKFELETHESYYPVSGFVAPHQYLSLLODRYFPIASVMRTLDKDNFSLTP 196  
DB 320 QRILNQLRNSCDFLGYRSVINTFVNRNQL-----ENFLDTRMKLLD--TTAPATP 374  
QY 197 DLIDLH-----LGH-VWMLHPSFSEFFI--NWRLFTYKTEKQVALPSKKORIQT 245  
DB 375 -LIGSLSKAAQCTGLGHPISKPIWTOLSDWDILDRFGICRNLFLH-YHSGSSKKQALYOL 432  
QY 246 QSNLTAIVRCFWFTVEGLIENHEGRKAYGAVLISSPQELGHAFIDNVRVLPLELDQIR 305  
DB 433 K-----YILR-----LSCARTLARKHK-----STVTFMQLRGLSVLEEFF--EEQVFS 476  
QY 306 LPFNTST-----PQETLFSIRHFDLV 327

DB 477 LMFARKTHFSGHSERIWFYDIIRIDDLV 507  
RESULT 41  
Q95F61 PRELIMINARY; PRT; 513 AA.  
AC Q95F61;  
DT 01-DEC-2001 (T-EMBLrel. 19, Created)  
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)  
DE Probable intron maturase (Maturase K).  
GN MATK  
OS Aegopogon cenchroides.  
OC Chloroplast  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
OC Chloridoideae; Chlorideae; Aegopogon.  
OX NCBI\_TaxID=160573;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hilu K.W., Alice L.A.;  
RT "A Phylogeny of Chloridoideae (Poaceae) Based on matk Sequences."  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II  
INTRONS (BY SIMILARITY).  
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,  
AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY  
MITOCHONDRIAL INTRONS.  
CC EMBL: AF312324; AAK60015.1; -;  
DR InterPro: IPR000442; Intron\_maturase2.  
DR InterPro: IPR002866; MatK\_N  
DR Pfam: PF01348; Intron\_maturase2; 1.  
DR Pfam: PF01824; MatK\_N; 1.  
KW Chloroplast; mRNA processing.  
SQ SEQUENCE 513 AA; 61640 MW; E80743D3EA97F086 CRC64;  
Query Match 5.2%; Score 99; DB 8; Length 513;  
Best Local Similarity 19.1%; Pred. No. 3.3;  
Matches 69; Conservative 63; Mismatches 124; Indels 106; Gaps 18;  
QY 22 QSLSLFFQNSQLO--AYSTPYSYRILLOKENKEQALARKHCISILEFFKNLLFV 77  
DB 199 KSIFLFSKRNKRLSRFLYNYFVSEYEFFLLRQKS-----SCLRTSYGTFLERI 249  
QY 78 HLLSLSKNREGCSTDMAVSTPFFNRNLWY-----LLSSRFSL-----W 118  
DB 250 HFRKMEH-----FGVMYFQFKTIWFLMEPLMHVRYQKVLASKGTLLPQKM 301  
QY 119 KSYCPRFFLDYLAFLGLSDPLDHLQAVIKFELETHESYYPVSGFVAPHQYLSLLODRYF 178  
DB 302 KSYIVNFSQYFFFFMAQPORICLNQUTNSCDFLGYRSNVPINTFLVTNQL-----ENFF 357  
QY 179 PIASVMRTLDKDNFSLTPDLIHLGHVPLIHPFSFSEFFINMGRFTK-----V 228  
DB 358 LIATRMKKFD--TTAPATP-LIRSL-----SKAQFCTGSGHPISKPIWTDLSWDI 405  
QY 229 IEKQAL-----PSKKORIOTLOSLNIAIVRCFWFTVEGLIENHEGRKAYGAVL 278  
DB 406 LDREGRICRNLFHYHSGSSKKQTLVRLK-----YILR-----LSCARTLARKHK 452  
QY 279 ISSPQELGHAFIDNVRVLPLELDQIRLPNTSTPQETLSIR--HFDLVELTSLKLEWM 336  
DB 453 RTFMQLRGLSVLEEFF--EEQVFSLMFS-----KTTYFSFHGHSERI-----WY 497  
QY 337 LD 338  
DB 498 LD 499  
RESULT 42  
Q95F52 PRELIMINARY; PRT; 513 AA.  
AC Q95F52;

DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-MAR-2002 (TREMblrel. 19, Last sequence update)  
 DE 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
 GN Probable Intron maturase (Maturase K).  
 OS Gymnospogon brevifolius.  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;  
 OC Chloridaceae; Chloridaceae; Gymnospogon.  
 ON NCBI\_TaxID=160579;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Hilu K.W., Alice L.A.;  
 RT "A Phylogeny of Chloridoideae (Poaceae) Based on matK Sequences."  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II  
 CC INTRONS (BY SIMILARITY).  
 CC -1- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,  
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY  
 CC MITOCHONDRIAL INTRONS.  
 DR EMBL: AF312333; AAK6024.1; -  
 DR InterPro: IPR000442; Intron\_maturase2.  
 DR Pfam: PF01348; Intron\_maturase2; 1.  
 DR Pfam: PF01824; MatK\_N; 1.  
 KM Chloroplast; mRNA processing.  
 SQ SEQUENCE 513 AA; 61432 MW; 56F4A4779579A0AF CRC64;

Query Match 5.2%; Score 99; DB 8; Length 513;

Best Local Similarity 19.3%; Pred. No. 3.3; 124; Indels 106; Gaps 18;  
 Matches 70; Conservative 62; Mismatches 124; Indels 106; Gaps 18;

QY 22 QSLSLFQNSQSLQRAVSTPYRYRIILQENKEKQALAHKICISILEFFKNLLFV 77  
 DB 199 KSIFLFSKKNKRLSFLVNSYVSEFEFLFLRKOS-----SCLNTSSGTF 246  
 QY 78 HLISLSKQREGCSTDMAVSTPFNNLMY-----RLSSRSFL-----W 118  
 DB 247 ERHIFSRKMEH-----FGVMYPGFEMKIWFEMDPLMHVYRGKYLILSKGTLKKKW 301  
 QY 119 KSYCPFFLDYLEAFGLSDPLDHOAVIKFELETHSYYPVAGFAPHOYLSILDRIYF 178  
 DB 302 KSYLVNFGYFSPFWTQGRIRLNQNTNSCFDGLGHSVPIITFLVRNQML-----ENFF 357  
 QY 179 PLASVWRITLDKNFSLTEPDLIHDLGHVPMILHSPSEPFIMNGRLFTK-----V 228  
 DB 358 LIDTRMKFD-TTAPAT-----LIGSL-----SKAQFCTSGHPIKPIYADLSMDI 405  
 QY 229 IEKVQAL-----PSKQRIQTLQSNLIAVRCFMTYVSGLIENHGRKAYGAVL 278  
 DB 406 LDFGRICNLFFHYHSSSKQOTLYRLK-----YILR-----LSCAKTLARKHK-----STV 452  
 QY 279 ISSPDLGAFIDNVRLPDLDOIIRLPNTSTPOETLFSIR--HFDELVELTSKLEWM 336  
 DB 453 RFFMRQLGSVFLSEFFT--BEDGVFSIMF-----PKTHSPFGSHSERI-----WY 497  
 QY 337 LD 338  
 DB 498 LD 499

RESULT 43

054921 PRELIMINARY; PRT; 924 AA.

ID 054921  
 AC 054921  
 DT 01-JUN-1998 (TREMblrel. 06, Created)  
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE Reeds  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OK NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RC MEDLINE=98070770; PubMed=9405631;  
 RA Kee Y., Yoo J.S., Hazuka C.D., Peterson K.E., Hsu S.C., Scheller R.H.;  
 RT "Subunit structure of the mammalian exocyst complex."  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:14438-14443(1997).  
 DR EMBL: AF032666; AAC01578.1; -  
 DR InterPro: IPR002909; IPT\_TIG.  
 DR Pfam: PF01833; TIG; 1.  
 SQ SEQUENCE 924 AA; 104031 MW; 1903C0593B113373 CRC64;

Query Match 5.2%; Score 99; DB 11; Length 924;

Best Local Similarity 19.9%; Pred. No. 6.7;  
 Matches 90; Conservative 77; Mismatches 122; Indels 164; Gaps 27;

QY 19 KLRQSLFQNSQSLQRAVSTPYRYRIILQENKEKQALAHKICISILEFFKNLLFV 78  
 DB 233 KVEGSMQKLEN--VNRASNTADTLFQEVIGRKD--ADSTRNALVILQRFK-FLFVL 286  
 QY 79 LLSLSKQREGCSTDMAVSTPFNNLMYRLSSR-----SLKSYCPFFLDYLEAF 133  
 DB 287 PLNKRNIQKQ--DYDVIND-----YKASLFGKTEVGVFKY-----YAEVB 329  
 QY 134 GLSDPLDHOAVIKFELETHSYYPVAGFV-----APHQY-LSLIQ 174  
 DB 330 AGIEDL--RELTLKLU-LTPSTLHDQKRYIRYISDLHAPGDPAWOCIGAHKMTLKLMO 386  
 QY 175 D-----RYPIASVWRITLDKNFSLTPDLIHDLGHV----- 206  
 DB 387 DCKGHNKSLKGNPGHSPW-----LDLDN-----DARPSVGHISQTRASLKGSSFGS 435  
 QY 207 ----PWLLH-----PSFSEPFIMN--GRLETKVIEK--VQALPSK 238  
 DB 436 GRDPTWRYKTPHRAVEKLTLYLSQLPNFMKIMISYVNGSLPSEFAEKSQIEKSKV 495  
 QY 239 KOR-----IQTLQSNLIAVRCFMTYVSGLIENHGRKAYGAVLIS--SPQELG 286  
 DB 496 RQRNDPKKMIQGVMSLVKYLIRGALLPFSIREG-----DGRQ-YGVMEVQAEISQWILA 549  
 QY 287 HAFIDNVV-----LPLEDOIIRLPNTSTPOETLFSIR-----HPEDEV 327  
 DB 550 HV-IQTRLYVESLTALBIPNDMLQI-----QDLILDRHICIMVTLQHTAEIK 599  
 QY 328 ELTSKLEWMTDGLSLEIPLVNOEKYLSGEVL 360  
 DB 600 RLAKEDMIYDNEGTLSPCFQEGSIVHSLSQSL 632

RESULT 44

08TPH6 PRELIMINARY; PRT; 1078 AA.

ID 08TPH6  
 AC 08TPH6  
 DT 01-JUN-2002 (TREMblrel. 21, Created)  
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Hypothetical protein MA1936.  
 GN MA1936.  
 OS Methanosarcina acetivorans.  
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;  
 OC Methanosarcinaceae; Methanosarcina.  
 ON NCBI\_TaxID=2214;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;  
 RC MEDLINE=21929760; PubMed=11932283;  
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,  
 RA Fitzhugh W., Calvo S., Engle R., Smirnov S., Annot D., Brown A.,  
 RA Allen L., Naylor J., Stenge-Thomann N., Dearellano K., Johnson R.,  
 RA Linton L., McEwan P., McKernan K., Talamas J., Tirell A., Ye W.,  
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grathme D.A., Guss A.M.,  
 RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,



```

QY 10 PKYILKIALKLRQSLSPFQNSOSLOPAYS-TPYSYRIILQXENKEKQALARKKICISIL 68
DB 29 PKFALFATLSSYKLSLFOKALYAOELFSLNPTSPNGIMLAKSYTEN--LRLEBANLNL 86
QY 69 -----EPPKNLIFVHLISLTKQRECGSTDMAVSTPPFNMLRYLLSRFS 116
DB 87 QTLTRKDDLEDELKELAFIYKLSNKLSESEQIFKEL--LSKDMYNNLN----- 133
QY 117 LKSYCPREFL-DYLEAFG-----LSDPLDQAVIKFELTFHSYVSGFVAPHQ 168
DB 134 LMKVAYEIFYGHDFTKALNHNHCHFMQDLID-----KLQIAEQ 177
QY 169 YLSU--LQDRYFPIASVWRTLDKONFSLT---PDLIHDLGHVPLLHPSP----- 214
DB 178 NLNLVTKLEDR-----LHSTKENLTIKIEDFLTHQILPQKAVLLFKLRISDSLFL 230
QY 215 -----SEFIMNGRLFTYIIE--KVQALPCKRQRIQTLQSNLIATVRCFPTVE 261
DB 231 OSLOEANOHAQFQWQ--YAKLEPNSYQDAHYAYKCSLDSHATYQPDLAYLMLR 266
QY 262 SGLIEN-HEGRKAYGAVLISSPOELGHAFIDNVRLPLE--LDQIIR--LPRNTSTPQ 314
DB 287 MGVQDNFEEGKKYYSRLFYAHNETPSTYHNSLKAFFKGVDAFKKKEVLPCEQOFG 346
QY 315 ETLFSIRHDELVELTSLKEMMLDQGLSIPLYNQKXLS 355
DB 347 DTIMYARCLERKCKIASKVLPAPQSAHYEMFK--NQIKFLN 385

```

## RESULT 47

OS94DZ7 PRELIMINARY; PRT; 945 AA.

```

ID OS94DZ7
AC 01-DEC-2001 (TREMBLREL. 19, Created)
DT 01-DEC-2001 (TREMBLREL. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLREL. 20, Last annotation update)
DE Putative zinc protease.
CN 00010810.10.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Euphorbiaceae; Oryzae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone: P0010810."
RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AP003224; BAB63566.1;
DR InterPro; IPR001431; Peptidase_M16.
DR Pfam; PF00675; Peptidase_M16; I.
DR PROSITE; PS00143; INSULINASE; UNKNOWN_1.
KM Protease.
SQ SEQUENCE 945 AA; 109356 MW; 0846F635508477F6 CRC64;

```

Query Match 5.2%; Score 98.5; DB 10; Length 945;

Best Local Similarity 22.8%; Pred. No. 7.6; Mismatches 75; Indels 123; Gaps 19;

```

QY 27 PFQNSOSLOPAYSSTPYRYRII-----LQENKEK----- 56
DB 442 FEGTDSVEPMYCTAIVSVENVTSMIQWIQKAPTEKLCIPKPNIFIPKDSLKAHEKV 501
QY 57 -QALARKKICISILEFFKNLIF-----VHL-----SLSKNREGSCTMAVSTPPFNRL 105
DB 502 KFPALIRKTPLSRLMYPMDLFTPKVHIVIDHCPCLTSHSE-----AVIST-----S 550
QY 106 LMYRLSSRFLMSKSYCPREFLDYLEAFG-----LSDPLDQAVIKFEL----- 151
DB 551 LEVDLLA-----DYLNAYGGYNDKMRILLDAIMKH-----ISNFVAVPMRNC 592

```

```

QY 152 ---ETHSYYPVSGFVAPH---OYLSU-LQDRYFPIASVWRTLDKONFSLTPLDLIDL 203
DB 593 ALKETAVKYOQNFKSQAPYQANSYLSLLEDDQWPMVEKLELKS---LEPD--SLA 645
QY 204 GHVPMLHPSPSEFFIMNGRL-----FTYKIEKVO-----ALPSKK--ORIQ 244
DB 646 KFPILHLSKTFELCYI-QGNIEPDATISVOIEIDTIFNTPKSVFMSPSQYLIRVIT 704
QY 245 LQSNLIATVRCF 256
DB 705 LENEL-----KCY 712

```

## RESULT 48

OS94DHL PRELIMINARY; PRT; 924 AA.

```

ID OS94DHL
AC 01-JUN-2001 (TREMBLREL. 17, Created)
DT 01-JUN-2001 (TREMBLREL. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLREL. 17, Last annotation update)
DE 2410030124Rik protein.
CN 2410030124Rik.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schiraldi L.M., Straubli P., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Oikido T., Furuno M., Aono H., Balderelli R., Barch G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bontaldo M.F.,
RA Brownstein M.J., Bull C., Flechner C., Fujita W., Gariboldi M.,
RA Gustinich S., Hill D., Hotmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima U., Mazzerelli U., Mombaert P.,
RA Nordone P., Ring B., Ringwald W., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
RA Wyshaw-Borls A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayaishizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK016532; BAB30290.1;
DR MGI; MGI:1913732; 2410030124Rik.
DR InterPro; IPR002909; IPT_TIG.
DR Pfam; PF01833; TIG; 1.
SQ SEQUENCE 924 AA; 103958 MW; 457BAD92BA3040B CRC64;

```

Query Match 5.2%; Score 98; DB 11; Length 924;

Best Local Similarity 20.0%; Pred. No. 8.1; Mismatches 125; Indels 160; Gaps 26;

```

QY 19 KLRQSLSPFQNSOSLOPAYSSTPYRYRIIQXENKEKQALARKKICISILEFFKNLIF 78
DB 233 KVBGSMTQKLEN--VLMRSNTADTLFOEVLGRQK--ADSTRNALNVQIRK-FLFNL 286
QY 79 LLSLSKNQREGSCTMAVSTPPFNRLMYRLSSRP-----SLMSYCPREFLDYLEAF 133
DB 287 PLNKRNQIGG--DYDVVIND-----YKASLSFGKTEVOYFKKY-----YAEVE 329
QY 134 GLSDPLDQAVIKFELTFHSYYPVSGFV-----APHQY-LSLQ 174
DB 330 AGIEDL--RELALKKL-LTPSTLHDQKRYRYSLDLHAGDDPAMQCIGAQHKMTKLQ 386
QY 175 D-----RYFPIASVWRTLDKONFSLTPLDLIDLGHV----- 206

```

```
Db 387 DCKEGHMKSLKGHPGPHSPM-----LDLDN-----DVRPSVLGHLQSQTASLKRGSFQS 435
QY 207 -----PWLLH-----PSFSEFFINM--GRFTKVKIEKV-QALPSK-- 238
Db 436 GRDDTWRYKTPHRVAFVEKLTKLVLSQLFNFKWLISYVNGSLFSETAEKSGQSERSKNV 495
QY 239 -----KORIQTQSNIILAIIVCFWFTVESGLIENHEGRKAYGAVLIS---SPQELGHA 288
Db 496 RQRONDFFKMIQEVWHSVLKIRGALLPLS---LREGDGRQ-YGGEVQVQALSGQWLAVH 551
QY 289 FIDNVRV-----LPLELDOIIRLPENTSTPOETLFSIR-----HFDELVEL 329
Db 552 -IQITRLTYESLTALEIPNDMLQII-----QDLIDIRICIMVTLOHTABEIKRL 601
QY 330 TSKLEWMLDQGLLESIPLYNQEKYLSGFVL 360
Db 602 ABKEDWVDVNEGLTSLPCQPEQSIHVHLSQSL 632

RESULT 49
Q9GZG6 PRELIMINARY; PRT; 1887 AA.
AC Q9GZG6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 214.2 kDa protein.
GN F56A6.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RC MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RC MEDLINE=99069613; PubMed=9851916;
RA Murray J., Rohlfing T., O'Neal D., Wilson R.;
RA "The sequence of C. elegans cosmid F56A6.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RC MEDLINE=99069613; PubMed=9851916;
RA Waterston R.;
RA "Direct Submission.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF067217; AAF99977.1; -.
DR HSSP; P08799; 1MND.
DR InterPro; IPR002219; DAG_PE-bind.
DR InterPro; IPR000048; IQ_Egion.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR000159; myosin_head.
DR InterPro; IPR000198; RhoGAP.
DR Pfam; PF00130; DAG_PE-bind; 2.
DR Pfam; PF00612; IQ; 4.
DR Pfam; PF00663; myosin_head; 2.
DR Pfam; PF00788; RA; 1.
DR Pfam; PF00620; RhoGAP; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00109; C1; 2.
DR SMART; SM00015; IQ; 4.
DR SMART; SM00242; MYSc; 1.
DR SMART; SM00314; RA; 1.
DR SMART; SM00324; RhoGAP; 1.
DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; UNKNOWN_2.
```

```
DR PROSITE; PS0081; DAG_PE_BIND_DOM_2; 2.
KW Hypothetical protein.
SQ SEQUENCE 1887 AA; 214207 MW; 992DDE7F409D2D84 CRC64;

Query Match
Best Local Similarity 23.3%; Score 98; DB 5; Length 1887;
Matches 71; Conservative 34; Mismatches 92; Indels 108; Gaps 16;

QY 9 DPKYILKALKRSLSPFQNSQSLQRAYSTP-----YSYRIILOKENKEQALA 60
Db 217 NPKY-----ARLYFQS-----KRLGSLPHIFAADVCHNMLRIKEN----- 254
QY 61 RHKICISI-----LEFFKNLAFVHLLSLSKNQREGCSTDMAVST-----PFNRLMY 108
Db 255 --QCVVSGSGSGKSTESTNHLMSHLISLSQKSGTGTLLSAGFVLEAFQNAVTLT 312
QY 109 RLLSRFSLWKSYPFRF-LDYLEAFGLLSDFLDHQVAKFPELET-----H 154
Db 313 NNNSSRFG-----KPIKINYREN-GWVSG-----ANVEIYLLEKSRIFQTKGERNYH 359
QY 155 FSYYPVSG-----FVAPHQYLSLLQDRYFPFIASVMTLKDKNFSLTPDLIHLLG 204
Db 360 VFYLLLEGADDEERKKYFLKPHDYKYLNQNEFFALEGVN-----ERNEF-----DRLR 408
QY 205 HYPMLLHPSFSEFFINMGRFLTQVIEKVOALPSKKQRIQTQLQSNLIAIVRCFWFTVESGL 264
Db 409 HA-----MSSVGFCAKQTQTFIGISAV-----LLGNITVTKRHGYHSDSGY 452
QY 265 IENHE 269
Db 453 IENEE 457

RESULT 50
Q9NJ03 PRELIMINARY; PRT; 239 AA.
AC Q9NJ03;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Tryptophan hydroxylase (Fragment).
OS Lymnaea stagnalis (Great pond snail).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;
OC Lymnaeidae; Lymnaea.
OX NCBI_TaxID=6523;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21359676; PubMed=11466431;
RA Koert C.E., Spencer G.E., van Minnen J., Li K.W., Geraerts W.P.,
RA Syed N.I., Smit A.B., van Kesteren R.E.;
RA "Functional implications of neurotransmitter expression during axonal
RT regeneration: serotonin, but not peptides, auto-regulate axon growth
RT of an identified central neuron.";
RL J. Neurosci. 21:5597-5606(2001).
DR EMBL; AF129815; AAF36488.1; -.
DR HSSP; P04176; 1PHZ.
DR InterPro; IPR001273; Aaa_hydroxylase.
DR Pfam; PF00351; biotin_H; 1.
DR PRINTS; PR00372; FWHYDEXLASE.
DR PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
DR NON_TER 1
FT NON_TER 239
SQ SEQUENCE 239 AA; 28001 MW; D721BD16C7D7B194 CRC64;

Query Match
Best Local Similarity 21.7%; Score 97.5; DB 5; Length 239;
Matches 34; Conservative 30; Mismatches 64; Indels 29; Gaps 5;

QY 71 FKNLFLVHLLSLSKNQREGCSTDMV-----VSTPFNRR-----NLWYRLLSRFRSLWKS 120
Db 93 FKDTVY-----RQRKIITLAKYAHGTEIFVEYTEVETWGTGIVFRELMLKLYPT 144
QY 121 YCPRFFLDYLEAFGLLSDFLDH-----QAVIKFFLETHFSYYPVSGFVAPHQYLSL 172
```

Fri Jan 10 10:56:14 2003

us-09-438-185a-1047.rspc

Page 24

Db 145 HACR--EYLANIPFLVEHCGYREDNVPQLEDISRFLKERTGFTLAFVAGYLSRDPLAG 201  
QY 173 LODRYFPPIASVMRTLDKDNFSLTPDLIHDLGHVPMI 209  
Db 202 LAFRVHCTQYIRHRDPFYTPEPPCCHLMGMPLL 238

Search completed: January 9, 2003, 16:59:53  
Job time : 90 secs





103	87.5	4.6	334	2	E71680	cytochrome C oxid
104	87.5	4.6	2	2	AH0469	probable 4-alpha-L
105	87.5	4.6	454	2	J00201	hypothetical 65K p
106	87.5	4.6	694	2	D81280	polysphosphate kina
107	87.5	4.6	1122	2	A97814	transcription-repa
108	87.5	4.6	1237	2	AC1583	intermalin protein
109	87.5	4.6	1817	2	H71611	probable secreted
110	87	4.6	293	2	F64558	conserved hypothet
111	87	4.6	324	1	A49953	choloalglycine hyd
112	87	4.6	741	2	H90507	atp-dependent heli
113	86.5	4.6	128	2	AH2322	hypothetical prote
114	86.5	4.6	533	2	T27589	hypothetical prote
115	86.5	4.6	710	2	B89252	protein ZC45.4 (l
116	86.5	4.6	990	2	S23416	lanthibiotic epide
117	86.5	4.6	1202	2	T20960	ribonuclease P (EC
118	86	4.6	339	2	A48773	hypothetical prote
119	86	4.6	422	2	A72309	conserved hypothet
120	86	4.6	512	2	D98074	conserved hypothet
121	86	4.6	565	2	D72222	conserved hypothet
122	86	4.6	1405	1	DJ2BA	DNA-directed DNA p
123	86	4.6	1822	2	S44849	K12H4.8 protein -
124	85.5	4.5	321	2	D90154	transposase ISC143
125	85.5	4.5	816	2	T25788	hypothetical prote
126	85.5	4.5	1489	2	T38842	probable RAS GTPas
127	85.5	4.5	1613	2	D90129	hypothetical prote
128	85.5	4.5	1808	2	AB1847	serine/threonine k
129	85.5	4.5	2039	2	S64540	probable calcium c
130	85	4.5	238	1	B64313	probable 3-isoprop
131	85	4.5	328	1	T32210	hypothetical prote
132	85	4.5	462	1	YSBYTM	threonine-CRM119
133	85	4.5	471	2	F71618	adenylonuclease 1
134	85	4.5	493	2	H95209	transcription acti
135	85	4.5	549	2	F64540	conserved hypothet
136	85	4.5	567	2	B63166	hypothetical prote
137	85	4.5	717	2	S78177	hypothetical prote
138	85	4.5	938	2	G70472	hypothetical prote
139	84.5	4.5	358	2	T22823	hypothetical prote
140	84.5	4.5	456	2	G71152	hypothetical prote
141	84.5	4.5	531	2	B84442	hypothetical prote
142	84.5	4.5	553	2	E95160	hypothetical prote
143	84.5	4.5	585	2	S48951	hypothetical prote
144	84.5	4.5	719	2	C88216	protein B0495.2 (l
145	84.5	4.5	820	2	T27909	hypothetical prote
146	84.5	4.5	1070	2	G84982	exodeoxyribonuclea
147	84.5	4.5	1107	1	S52517	myosin I heavy cha
148	84.5	4.5	1748	1	JQ1555	genome polyprotein
149	84.5	4.5	3068	1	A44062	conserved hypothet
150	84	4.4	240	2	F70409	

## ALIGNMENTS

RESULT 1  
E72002  
aromatic amino acid hydroxylase - Chlamydomonas pneumoniae (strain CML029)  
C:Species: Chlamydomonas pneumoniae, Chlamydia pneumoniae  
C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000  
C:Accession: E72002  
R:Kelman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:92060606; PMID:10192388  
A:Accession: E72002  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-362 <ARN>  
A:Cross-references: GB:AE001665; GB:AE001363; NID:g4377376; PIDN:AAD19183.1; PID:g437738  
A:Experimental source: strain CML029  
C:Genetics:  
A:Gene: CP1046

Query Match 99.8%; Score 1886; DB 2; Length 362;

Best Local Similarity 99.7%; Pred. No. 2e-144; Matches 361; Conservative 1; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	VHNCERTLPKYLILKALKRQSLSPFQNSQSLQRYSTPVSYYRIILQENKREKOLA	60						
Db	1	MHNCERTLPKYLILKALKRQSLSPFQNSQSLQRYSTPVSYYRIILQENKREKOLA	60						
Qy	61	RHKICISLEFFPKNLLFVHLISLSKQRCGCTDAVAVSTFFRNLMYRLSSRSFLWS	120						
Db	61	RHKICISLEFFPKNLLFVHLISLSKQRCGCTDAVAVSTFFRNLMYRLSSRSFLWS	120						
Qy	121	YCRPFLDYLEAFGLSDPLDQAVIKFPELETHFSYYPVSGFVADHGYLSLQDRYFPI	180						
Db	121	YCRPFLDYLEAFGLSDPLDQAVIKFPELETHFSYYPVSGFVADHGYLSLQDRYFPI	180						
Qy	181	ASVWRITLDKONSLTPDILHDLGHVPMILHPSSEFFIMGRLEFKVIEKVALPSKKQ	240						
Db	181	ASVWRITLDKONSLTPDILHDLGHVPMILHPSSEFFIMGRLEFKVIEKVALPSKKQ	240						
Qy	241	RIOTLOSNIATVRCFMTVESGLIENHGRKAGAVLISPOELGHAFIDNVRVPLEL	300						
Db	241	RIOTLOSNIATVRCFMTVESGLIENHGRKAGAVLISPOELGHAFIDNVRVPLEL	300						
Qy	301	DOIIIRLPMTSTPQETLSIRHPDELVELTSKLEMDQGLSIFLYNQEKYLSGFVYL	360						
Db	301	DOIIIRLPMTSTPQETLSIRHPDELVELTSKLEMDQGLSIFLYNQEKYLSGFVYL	360						
Qy	361	CO 362							
Db	361	CO 362							
RESULT 2									
aromatic amino acid hydroxylase (imported) - Chlamydomonas pneumoniae (strain J138)									
C:Species: Chlamydomonas pneumoniae, Chlamydia pneumoniae									
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001									
C:Accession: C86621									
R:Shiwa, M.; Hirakawa, H.; Kinoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiwa, T.									
Nucleic Acids Res. 28, 2311-2314, 2000									
A:Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.									
A:Reference number: A86491; MUID:20330349; PMID:10871362									
A:Accession: C86621									
A:Status: preliminary									
A:Molecule type: DNA									
A:Residues: 1-362 <STO>									
A:Cross-references: GB:BA000008; NID:g8979419; PIDN:BA09253.1; GSPDB:GN00142									
A:Experimental source: strain J138									
C:Genetics:									
A:Gene: CP1046									
Query Match 99.8%; Score 1886; DB 2; Length 362;									
Best Local Similarity 99.7%; Pred. No. 2e-144; Matches 361; Conservative 1; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	VHNCERTLPKYLILKALKRQSLSPFQNSQSLQRYSTPVSYYRIILQENKREKOLA	60						
Db	1	MHNCERTLPKYLILKALKRQSLSPFQNSQSLQRYSTPVSYYRIILQENKREKOLA	60						
Qy	61	RHKICISLEFFPKNLLFVHLISLSKQRCGCTDAVAVSTFFRNLMYRLSSRSFLWS	120						
Db	61	RHKICISLEFFPKNLLFVHLISLSKQRCGCTDAVAVSTFFRNLMYRLSSRSFLWS	120						
Qy	121	YCRPFLDYLEAFGLSDPLDQAVIKFPELETHFSYYPVSGFVADHGYLSLQDRYFPI	180						
Db	121	YCRPFLDYLEAFGLSDPLDQAVIKFPELETHFSYYPVSGFVADHGYLSLQDRYFPI	180						
Qy	181	ASVWRITLDKONSLTPDILHDLGHVPMILHPSSEFFIMGRLEFKVIEKVALPSKKQ	240						
Db	181	ASVWRITLDKONSLTPDILHDLGHVPMILHPSSEFFIMGRLEFKVIEKVALPSKKQ	240						
Qy	241	RIOTLOSNIATVRCFMTVESGLIENHGRKAGAVLISPOELGHAFIDNVRVPLEL	300						
Db	241	RIOTLOSNIATVRCFMTVESGLIENHGRKAGAVLISPOELGHAFIDNVRVPLEL	300						
Qy	301	DOIIIRLPMTSTPQETLSIRHPDELVELTSKLEMDQGLSIFLYNQEKYLSGFVYL	360						
Db	301	DOIIIRLPMTSTPQETLSIRHPDELVELTSKLEMDQGLSIFLYNQEKYLSGFVYL	360						
Qy	361	CO 362							
Db	361	CO 362							

DB 241 RIOTLQNLIAIVRCFWFTVESGLINHEGRKAYGAVLISSPQELGHAFIDNVRVLPLEL 300  
QY 301 DOIIRLPNTSTPQELTFSIRHFDLVELTSLKLEMLDQGLLESIPLYNQEKYLSGFVFL 360  
DB 301 DOIIRLPNTSTPQELTFSIRHFDLVELTSLKLEMLDQGLLESIPLYNQEKYLSGFVFL 360  
QY 361 CQ 362  
DB 361 CQ 362

RESULT 3  
phenylalanine-4-hydroxylase VCA0828 [imported] - Vibrio cholerae (strain N16961 serogroup O1, biotype El Tor)  
C:Species: Vibrio cholerae  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: D82413  
R:Heidelberger, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.; Chaudson, D.; Ermlaeva, W.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, P.; R.K.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
A:Title: DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.  
A:Reference number: A82035; MUID:20406833; PMID:10952301  
A:Accession: D82413  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-289 <HEI>  
A:Cross-references: GB:AE004410; GB:AE003853; NID:99658244; PIDN:AAF96726.1; GSPDB:GN001  
A:Experimental source: serogroup O1, strain N16961, biotype El Tor  
C:Genetics:  
A:Gene: VCA0828  
A:Map position: 2

Query Match 13.6%; Score 256.5; DB 2; Length 289;  
Best Local Similarity 25.6%; Pred. No. 3.2e-13;  
Matches 65; Conservative 56; Mismatches 108; Indels 25; Gaps 6;

QY 106 LWYLLSRFLSKSYCPREFDYLLEAFGLLSDFLDHQAQV-KFELETHFSYYPVSGFV 164  
DB 49 VHELTIRQEEVTRACQAYLDGLMLPDLRPPQPEINVLQRETNQVQVPPALI 108  
QY 165 APHQVLSLLQDRVPPIASVMTLDKDNFSLTDPDLHDLGHVPMLLHPSFSEFFINMRL 224  
DB 109 SDRFALLADKFPVATFLRRBEEDFYLQEPDFHEVYGHCAMLTHPDFAAFTHVYGL 168  
QY 225 FTKVLEKQALPSKKQRIQTLQNLIAIVRCFWFTVESGLINHEGRKAYGAVLISSPQE 284  
DB 169 GAKATPKERSY-----LARYWFTVEFGLVQEQGQTKYGGGILSSPGE 212  
QY 285 LGHAFIDNV-RVLPLELDQIIRLPNTSTPQELTFSIRHFDLVELTSLK-----LEWM-LD 338  
DB 213 TLVASESTIPKREFDIMQVLRTPYRIDIMQPIYVYVLPDLSQLYOLSORDVMALVWQMQ 272  
QY 339 QGLLESIPLYNQEK 352  
DB 273 DGLLP--PLFQPK 284

RESULT 4  
phenylalanine-4-hydroxylase PA0872 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: F83535  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Berman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.  
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen.  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: F83535  
A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-262 <STO>  
A:Cross-references: GB:AE004522; GB:AE004091; NID:99946768; PIDN:AAG04261.1; GSPDB:GN001;  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: phbA; PA0872

Query Match 12.5%; Score 235.5; DB 2; Length 262;  
Best Local Similarity 24.5%; Pred. No. 1.4e-11;  
Matches 61; Conservative 52; Mismatches 103; Indels 33; Gaps 6;

QY 93 DMVVSTPFFNRNLWYLLSRFLSKSYCPREFDYLLEAFGLLSDFLDHQAQVIRKFELE 152  
DB 13 DNGFIHYPETEHQVMTLITRQLKVIEGRAQCEYLDGIEQLG-----LPHERIPQLDEIN 67  
QY 153 -----THFSYYPVSGFVAPHOYLSILODRYFPIASVMTLDKDNFSLTDPDLHDLGHV 206  
DB 68 RVLAQTTCWRVAVPALIPFQTFELLASQFPVATIRTPPELDYLOEPDIFHEIFGHC 127  
QY 207 PWLHPSFSEFFINMRLTQVIEKQVQLPSKKQRIQTLQNLIAIVRCFWFTVESGLIE 266  
DB 128 PLLNPFPAETHTYKGLKA-----SKEER-----VFLARLYWMTIEFGLVE 171  
QY 267 NHEGRKAYGAVLISSPQELGHAFIDNV-RVLPLELDQIIRLPNTSTPQELTFSIRHF 323  
DB 172 TDQGRYGGGILSSPKETVYSLSDPELHQAQVLE---AMRTPYRIDILQPLVFLPDL 228  
QY 324 DELVELTSK 332  
DB 229 KRLFLAQE 237

RESULT 5  
S10489  
tryptophan 5-monoxygenase (EC 1.14.16.4) - human  
N:Alternate names: tryptophan 5-hydroxylase  
C:Species: Homo sapiens (nan)  
C>Date: 04-Sep-1998 #sequence\_revision 04-Sep-1998 #text\_change 03-Mar-2000  
C:Accession: S10489; S51559  
R:Soularand, S.; Darmon, M.C.; Ganem, Y.; Launay, J.M.; Mallet, J.  
Nucleic Acids Res. 18, 4257, 1990  
A:Title: Complete coding sequence of human tryptophan hydroxylase.  
A:Reference number: S10489; MUID:90332431; PMID:2377472  
A:Accession: S10489  
A:Molecule type: mRNA  
A:Residues: 1-444 <BOU>  
A:Cross-references: EMBL:X52836; NID:937954; PIDN:CAA37018.1; PID:G37955  
R:Tipper, J.P.; Citron, B.A.; Ribeiro, P.; Kaufman, S.  
Arch. Biochem. Biophys. 315, 445-453, 1994  
A:Title: Cloning and expression of rabbit and human brain tryptophan hydroxylase cDNA in  
A:Reference number: S51559; MUID:95077422; PMID:7986090  
A:Accession: S51559  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-18, 'T', '20-67, 'T', '69-89, 'TP', '92-96, 'M', '98-99, 'E', '101-103, 'S', '105-150, 'S', '153', 'G', '437-444 <TP>  
A:Cross-references: GB:L29306; NID:9531192; PIDN:AAA67050.1; PID:9531193  
C:Genetics:  
A:Gene: GDB:TPH; TPH  
A:Cross-references: GDB:L20732; OMIM:191060  
A:Map position: lip5.1-lip4.3  
C:Function:  
A:Description: catalyzes the oxidation of tryptophan to 5'-hydroxytryptophan by tetrahyd;  
A:Pathway: melatonin biosynthesis; serotonin biosynthesis; tryptophan catabolism  
C:Superfamily: phenylalanine 4-monoxygenase  
A:Keywords: bioprotein; iron; melatonin biosynthesis; metalloprotein; monooxygenase; oxid;  
F:58/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status predicted  
F:260,443/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #stat  
F:272,277,317/Binding site: iron (His, His, Glu) #status predicted

Query Match 12.3%; Score 231.5; DB 1; Length 444;  
Best Local Similarity 27.8%; Pred. No. 5.8e-11;  
Matches 64; Conservative 43; Mismatches 96; Indels 27; Gaps 4;



DB 343 QDIG-----LASLGATDEIEKLAFL-----YWFVEGLCRQNGIVKAYGAGL 386  
QY 279 ISSQPELGHAFIDNVRLVPLELQIIRLPFNFTSTPQETLFSIRHFDE 325  
DB 387 LSSYGELIHSLSDEPEVRDPDPAAAVQPCDQYQPVYFVSESFS 433

## RESULT 9

## WRTW

tryptophan 5-monooxygenase (EC 1.14.16.4) - rat  
N:Alternate names: tryptophan 5-hydroxylase  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 03-Mar-2000  
C:Accession: J00034; A60034; A24367  
R:Damon, M.C.; Guilbert, B.; Leviel, V.; Ehret, M.; Maitre, M.; Mallet, J.  
J. Neurochem. 51, 312-316, 1988  
A:Title: Sequence of two mRNAs encoding active rat tryptophan hydroxylase.  
A:Reference number: J00034; MUID:88244702; PMID:3379411  
A:Accession: J00034  
A:Molecule type: mRNA  
A:Residues: 1-444 <DAR>  
A:Cross-references: GB:X53501; NID:g57760; PIDN:CAA37579.1; PID:g57761  
A:Experimental source: pineal gland  
R:Kim, K.S.; Wessel, T.C.; Stone, D.M.; Carver, C.H.; Joh, T.H.; Park, D.H.  
Brain Res. Mol. Brain Res. 9, 277-283, 1991  
A:Title: Molecular cloning and characterization of cDNA encoding tryptophan hydroxylase  
A:Reference number: A60034; MUID:91245924; PMID:1645430  
A:Accession: A60034  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-444 <KIM>  
A:Experimental source: dorsal raphe nucleus  
R:Damon, M.C.; Grims, B.; Cash, C.D.; Maitre, M.; Mallet, J.  
FEBS Lett. 206, 43-46, 1986  
A:Title: Isolation of a rat pineal gland cDNA clone homologous to tyrosine and phenylalanine hydroxylase  
A:Reference number: A24367; MUID:87005247; PMID:2875901  
A:Accession: A24367  
A:Molecule type: mRNA  
A:Residues: 167-261 <DA2>  
A:Cross-references: GB:M28000; NID:g207432; PIDN:AAA42262.1; PID:g207433  
C:Comment: This enzyme has different physical properties in pineal gland and in dorsal raphe nucleus  
C:Function:  
A:Pathway: catalyzes the oxidation of tryptophan to 5'-hydroxytryptophan by tetrahydrobiopterin, melatonin biosynthesis; serotonin biosynthesis; tryptophan catabolism  
C:Superfamily: phenylalanine 4-monooxygenase  
C:Keywords: bipterin; iron; melatonin biosynthesis; metalloprotein; monooxygenase; oxidoreductase; tyrosine hydroxylase  
F:58/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status predicted  
F:260,443/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #status predicted  
F:272,277,317/Binding site: iron (His, His, Glu) #status predicted

Query Match 12.0%; Score 227.5; DB 1; Length 444;  
Best Local Similarity 27.4%; Pred. No. 1.2e-10;  
Matches 63; Conservative 43; Mismatches 97; Indels 27; Gaps 4;  
QY 107 WYRLSSRFLSKSYCRFFLDYLEAFGLLSDFLDH-----QAVIKFFELETHFSY 158  
DB 174 WGTIFRELNLKLYPTHACR---EYLRNLPLLSKYCYREDNTPQLEDVSNFLKERTGFSIR 230  
QY 159 PVSGFVAPHOYVLSLQDRYFPPIASVNRITLDKDNFSLTPDLIHLGHVPMWLLHPSFSEFF 218  
DB 231 PVAGYLSPRDFLSGLAFRVFCHTQYVRHSSDPLYTPEPTCHELLGHVPLLAEPSPAQFS 290  
QY 219 INMGRLTKVIEKVOALPSKKORIOTLOSNIATVRCFWFTVSGLIENHEGRKAYGAVL 278  
DB 291 QBIG-----LASLGASEETVQKLAT-----CYFFTFVEGLCKDQGLRVFGAGL 334  
QY 279 ISSQPELGHAFIDNVRLVPLELQIIRLPFNFTSTPQETLFSIRHFDELVE 328  
DB 335 LSSISELHSLHSGHAKVKPDPKPVACKOECLITSFDVYFVSESFEDEAKE 384

## RESULT 10

tryptophan 5-monooxygenase (EC 1.14.16.4) - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 22-Jun-1990 #sequence\_revision 09-Oct-1992 #text\_change 31-Mar-2000  
C:Accession: A34582  
R:Stoll, J.; Kozak, C.A.; Goldman, D.  
Genomics 7, 88-96, 1990  
A:Title: Characterization and chromosomal mapping of a cDNA encoding tryptophan hydroxylase  
A:Reference number: A34582; MUID:90243261; PMID:2110547  
A:Accession: A34582  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-447 <STO>  
A:Cross-references: GB:J04758; NID:g202113; PIDN:AAA63401.1; PID:g202114  
A:Note: the authors translated the codon AAC for residue 405 as Gln  
C:Superfamily: phenylalanine 4-monooxygenase  
C:Keywords: bipterin; iron; metalloprotein; oxidoreductase; phosphoprotein  
F:275,280,320/Binding site: iron (His, His, Glu) #status predicted  
Query Match 12.0%; Score 226.5; DB 2; Length 447;  
Best Local Similarity 27.4%; Pred. No. 1.5e-10;  
Matches 63; Conservative 43; Mismatches 97; Indels 27; Gaps 4;  
QY 107 WYRLSSRFLSKSYCRFFLDYLEAFGLLSDFLDH-----QAVIKFFELETHFSY 158  
DB 177 WGTIFRELNLKLYPTHACR---EYLRNLPLLSKYCYREDNTPQLEDVSNFLKERTGFSIR 233  
QY 159 PVSGFVAPHOYVLSLQDRYFPPIASVNRITLDKDNFSLTPDLIHLGHVPMWLLHPSFSEFF 218  
DB 234 PVAGYLSPRDFLSGLAFRVFCHTQYVRHSSDPLYTPEPTCHELLGHVPLLAEPSPAQFS 293  
QY 219 INMGRLTKVIEKVOALPSKKORIOTLOSNIATVRCFWFTVSGLIENHEGRKAYGAVL 278  
DB 294 QBIG-----LASLGASEETVQKLAT-----CYFFTFVEGLCKDQGLRVFGAGL 337  
QY 279 ISSQPELGHAFIDNVRLVPLELQIIRLPFNFTSTPQETLFSIRHFDELVE 328  
DB 338 LSSISELHSLHSGHAKVKPDPKPVACKOECLITSFDVYFVSESFEDEAKE 387

## RESULT 11

## JN0068

tyrosine 3-monooxygenase (EC 1.14.16.2) - mouse  
N:Alternate names: tyrosine hydroxylase  
C:Species: Mus musculus (house mouse)  
C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 31-Mar-2000  
C:Accession: JN0068; S21322  
R:Ichikawa, S.; Sasakawa, T.; Nagatsu, T.  
Biochem. Biophys. Res. Commun. 176, 1610-1616, 1991  
A:Title: Primary structure of mouse tyrosine hydroxylase deduced from its cDNA.  
A:Reference number: JN0068; MUID:91248263; PMID:1674869  
A:Accession: JN0068  
A:Molecule type: mRNA  
A:Residues: 1-498 <ICH>  
A:Cross-references: GB:M69200; NID:g201997; PIDN:AAA40434.1; PID:g201998  
A:Experimental source: brain  
R:Morgan, W.W.; Bermudez, J.; Sharp, Z.D.  
submitted to the EMBL Data Library, June 1990  
A:Description: DC-12 Nuclear Extracts Produce Tissue-Specific Protection of Several Sequences from Oxidative Damage  
A:Reference number: S21322  
A:Accession: S21322  
A:Molecule type: DNA  
A:Residues: 1-30 <MOR>  
A:Cross-references: EMBL:X53503; NID:g55055; PIDN:CAA37580.1; PID:g55056  
C:Comment: This enzyme, which requires ferrous iron, catalyzes the hydroxylation of tyrosine in the physiology of adrenergic neurones.  
C:Superfamily: phenylalanine 4-monooxygenase  
C:Keywords: bipterin; iron; metalloprotein; monooxygenase; oxidoreductase; phosphoprotein  
F:331,336,376/Binding site: iron (His, His, Glu) #status predicted  
Query Match 11.9%; Score 225.5; DB 2; Length 498;  
Best Local Similarity 27.5%; Pred. No. 2.1e-10;  
Matches 66; Conservative 42; Mismatches 101; Indels 31; Gaps 5;

Qy 107 WYRLSSRFSLMKSVCRFPLDYIEAFGLISDFLDH-----QAVIKPELETHNSY 158  
 Db 233 MEKVATKLGKATPHACK--EHLAEQALERKCGREDSTPOLEEVSHFKERKQCLR 289  
 Qy 159 PVSGFVAAPHQYLSLQDRYFPFASVMTLKDKNFSLTPDLIHDLGHVPMILHPSSEFP 218  
 Db 290 PVAGLLSARDPLASLAFVFCQCYIHHASSPMHSEPPCCHELLGHVPMILADRTFAQFS 349  
 Qy 219 IMNGRLFTKVIKQVQLPSKKQRIQTLQSNLAIIVRCFMTVESGLIENHGRKAYGAVL 278  
 Db 350 QDIG-----LASLGASDEELBKSLST-----VWFTEVFGICKONELKAYGAGL 393  
 Qy 279 ISSPOLGHAFLDNVRLPLDLQIIRLPENTSTPOETLPSIRHF-----DELVELTSKLE 334  
 Db 394 LSSVGEHLHSLSEEPVRAFPDPAVQYQDTQYQVYFVSESFSADAKDLRNYSRIQ 453

## RESULT 12

151567  
 tryptophan 5-monooxygenase (EC 1.14.16.4) - African clawed frog  
 C:Species: Xenopus laevis (African clawed frog)  
 C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 31-Mar-2000  
 C:Accession: I51567  
 R:Green, C.B.; Besharar, J.C.  
 J. Neurochem. 62, 2420-2428, 1994  
 A:Title: Tryptophan hydroxylase expression is regulated by a circadian clock in Xenopus  
 A:Reference number: I51567; MUID:94246419; PMID:8189245  
 A:Accession: I51567  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-481 <GRE>  
 A:Cross-references: GB:L20679; NID:9450644; PIDN:AAA21306.1; PID:9450645  
 C:Superfamily: phenylalanine 4-monooxygenase  
 C:Keywords: bioprotein, iron, metalloprotein, oxidoreductase  
 F:309,314,354/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.8%; Score 223.5; DB 2; Length 481;  
 Best Local Similarity 29.5%; Pred. No. 2.9e-10;  
 Matches 70; Conservative 38; Mismatches 90; Indels 39; Gaps 9;  
 Qy 107 WYRLSSRFSLMKSVCRFPLDYIEAFGLISDFLDH-----QAVIKPELETHNSY 158  
 Db 211 WQTVREPLANKYPTHACK--EYKULPLSKKCGIREDNIOLEEDVSFLKERTGFTIR 267  
 Qy 159 PVSGFVAAPHQYLSLQDRYFPFASVMTLKDKNFSL--TDGLIHDLGHVPMILHPSSEFP 217  
 Db 268 PVAGLLSARDPLASLAFVFCQCYIHHASSPMHSEPPCCHELLGHVPMILADRTFAQFS 346  
 Qy 219 IMNGRLFTKVIKQVQLPSKKQRIQTLQSNLAIIVRCFMTVESGLIENHGRKAYGAVL 277  
 Db 327 SQEIG-----LASLGASDEAVQKLAFL-----CYFTEVFGICKONELKAYGAGL 370  
 Qy 279 ISSPOLGHAFLDNVRLPLDLQIIRLPENTSTPOETLPSIRHF-----DELVELTSKLE 334  
 Db 394 LSSVGEHLHSLSEEPVRAFPDPAVQYQDTQYQVYFVSESFSADAKDLRNYSRIQ 453

## RESULT 13

WHRTY  
 tyrosine 3-monooxygenase (EC 1.14.16.2) - rat  
 N:Alternate names: tyrosine 3-hydroxylase  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 03-Mar-2000  
 C:Accession: A00510; A44714; S03026; I58264  
 R:Germa, B.; Lamouroux, A.; Blano, F.; Faucon Biguet, N.; Mallet, J.  
 Proc. Natl. Acad. Sci. U.S.A. 82, 617-621, 1985  
 A:Title: Complete coding sequence of rat tyrosine hydroxylase mRNA.  
 A:Reference number: A00510; MUID:85113249; PMID:2857492  
 A:Accession: A00510  
 A:Molecule type: mRNA  
 A:Residues: 1-498 <GR1>  
 A:Cross-references: GB:M10244; NID:g207408; PIDN:AAA42257.1; PID:g207409

R:Campbell, D.G.; Hardie, D.G.; Walliet, P.R.  
 J. Biol. Chem. 261, 10489-10492, 1986  
 A:Title: Identification of four phosphorylation sites in the N-terminal region of ty  
 A:Reference number: A44714; MUID:86278113; PMID:2874140  
 A:Accession: A44714  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 2-12,16-24,38-47,151-157 <CAN>  
 R:Bonnefoy, E.; Ferrara, P.; Rohrer, H.; Gros, F.; Tribault, J.  
 Eur. J. Biochem. 174, 685-690, 1988  
 A:Title: Role of the N-terminus of rat pheochromocytoma tyrosine hydroxylase in the  
 A:Reference number: S03026; MUID:88271342; PMID:2899026  
 A:Accession: S03026  
 A:Molecule type: protein  
 A:Residues: 2-26 <BON>  
 R:Harrington, C.A.; Lewis, E.J.; Krzemien, D.; Chikaraishi, D.M.  
 Nucleic Acids Res. 15, 2363-2384, 1987  
 A:Title: Identification and cell type specificity of the tyrosine hydroxylase gene p  
 A:Reference number: I58264; MUID:8174758; PMID:2882469  
 A:Accession: I58264  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-30 <RBS>  
 A:Cross-references: EMBL:X04914; NID:g57355; PIDN:CAA28584.1; PID:g57356  
 C:Function:  
 A:Description: catalyzes the 3'-hydroxylation of tyrosine to 3',4'-dihydroxyphenylal  
 A:Pathway: catecholamine biosynthesis  
 A>Note: this is the rate-limiting step in catecholamine biosynthesis  
 C:Superfamily: phenylalanine 4-monooxygenase  
 C:Keywords: bioprotein; catecholamine biosynthesis; iron; metalloprotein; monooxyg  
 F:8/Binding site: phosphate (Ser) (covalent) (by unidentified kinase) #status experi  
 F:19/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #stat  
 F:40,133/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #statu  
 F:331,336,376/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.8%; Score 223; DB 1; Length 498;  
 Best Local Similarity 26.7%; Pred. No. 3.3e-10;  
 Matches 65; Conservative 39; Mismatches 103; Indels 36; Gaps 5;  
 Qy 112 SSRPSFLMKS-----YCRPFUDYIEAFGLISDFLDH-----QAVIKPELETHNF 155  
 Db 227 ABEATKKEVYVTKGLATYTHACKENHLEGLRKYRGYEDSIPOLEEDVSFLKERTCF 286  
 Qy 156 SYTVSGFVAAPHQYLSLQDRYFPFASVMTLKDKNFSLTPDLIHDLGHVPMILHPSFS 215  
 Db 287 QLRVAGLLSARDPLASLAFVFCQCYIHHASSPMHSEPPCCHELLGHVPMILADRTFA 346  
 Qy 216 EFINNGRLFTKVIKQVQLPSKKQRIQTLQSNLAIIVRCFMTVESGLIENHGRKAYG 275  
 Db 347 QFQODIG-----LASLGASDEELBKSLST-----VWFTEVFGICKONELKAYG 390  
 Qy 276 AVLISSPOLGHAFLDNVRLPLDLQIIRLPENTSTPOETLPSIRHF-----DELVELTS 331  
 Db 391 AGLLSSVGEHLHSLSEEPVRAFPDPAVQYQDTQYQVYFVSESFSADAKDLRNYS 450  
 Qy 332 KLE 334  
 Db 451 RIQ 453

## RESULT 14

145983  
 tyrosine 3-monooxygenase (EC 1.14.16.2) - bovine  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 31-Mar-2000  
 C:Accession: I45983  
 R:DeWello, S.R.; Weisberg, E.P.; Stachowiak, M.K.; Turzai, L.M.; Gioto, A.E.; Kaplan,  
 J. Neurosci. Res. 19, 440-449, 1988  
 A:Title: Isolation and nucleotide sequence of a cDNA clone encoding bovine adrenal ty  
 A:Reference number: I45983; MUID:88259287; PMID:2898537  
 A:Accession: I45983  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA

A;Accession: JE0014  
A;Molecule type: DNA  
A;Residues: 1-30,35-135 <KOB3>  
A;Cross-references: GB:D00269; NID:g220099; PIDN:BAA25095.1; PID:g2951765  
C;Superfamily: phenylalanine 4-monoxygenase  
C;Keywords: biotin; iron; metalloprotein; monooxygenase; oxidoreductase  
F:324,329,369/Binding site: iron (His, His, Glu) \$status Predicted

Query Match 11.7%; Score 221.5; DB 2; Length 491;  
Best Local Similarity 27.1%; Pred. No. 4.3e-10;  
Matches 65; Conservative 42; Mismatches 102; Indels 31; Gaps 5;

QY 107 WYLLSSRFSLMWSYCPFRFLDYLAFAGLSDFLDH-----QAVIKFPELETHFSYV 158  
DB 226 WKEYVSTLGLYPTHACR---EHLEAFELLERFCGYREDRIPOLEDVSRLKERTGFQLR 282

QY 159 PVSFGVAPHOYLSLODRVFPPTASVMRTLDKDNLSLPDLIHLLGHVPHLLHPFSSEFF 218  
DB 283 PAAGLLSARDFLASLAFRVQCTQYIRHASSPMSPPECCHELLGHVPLADRTFAQFS 342

QY 219 INMGRLTKVIEQVQALPSKKORIOTLOSNIIVRCFWPTVSGLIENHGRKAYGAVL 278  
DB 343 QDIG-----LASLGVSDBEIKLST-----LYWFTVFGCLKQNGEVKAYGAGL 386

QY 279 ISSPOELGHAFIDNVRLPLELDQIIRLPNTSTPQETLSIRHF-----DELVELTSKLE 334  
DB 387 LSSYGELLHSLSEPEIRAFDPDAAAVQPDQYQYVYFVSFSFSDAKLRSYASRIQ 446

RESULT 15  
WHUHV4  
Tyrosine 3-monoxygenase (EC 1.14.14.16.2), splice form 4 - human  
N;Alternate names: tyrosine 3-hydroxylase  
N;Contains: tyrosine 3-monoxygenase, splice form 1; tyrosine 3-monoxygenase, splice fo  
C;Species: Homo sapiens (man)  
C;Date: 31-Mar-1989 #sequence revision 31-Mar-1989 #text change 16-Jun-2000  
C;Accession: A30002; A46825; A60201; JE0012; JE0013; JE0014; A27791; B27791; PNC  
R;Nagata, T.  
submitted to GenBank, December 1987  
A;Reference number: A94509  
A;Accession: A30002  
A;Molecule type: mRNA  
A;Residues: 1-528 <NAG1>  
A;Cross-references: GB:M17589; NID:g339680; PIDN:AAA61179.1; PID:g339681  
R;Kameda, N.; Kobayashi, K.; Ichinose, H.; Kishi, F.; Nakazawa, A.; Kurosawa, Y.; Fujita  
Biochem. Biophys. Res. Commun. 146, 971-975, 1987  
A;Title: Isolation of a novel cDNA clone for human tyrosine hydroxylase: alternative RNA  
A;Reference number: A90136; MUID:87298614; PMID:2887169  
A;Accession: A26825  
A;Molecule type: mRNA  
A;Residues: 1-94 <NAG2>  
A;Cross-references: GB:M17589; NID:g339680; PIDN:AAA61179.1; PID:g339681  
R;Le Bourdelle, B.; Boularand, S.; Boni, C.; Horellou, P.; Dumas, S.; Grima, B.; Mallet  
J. Neurochem. 50, 988-991, 1988  
A;Title: Analysis of the 5' region of the human tyrosine hydroxylase gene: combinatorial  
A;Reference number: A60201; MUID:88117543; PMID:2892893  
A;Accession: A60201  
A;Molecule type: mRNA  
A;Residues: 1-65 <LEB>  
A;Cross-references: GB:M24790; NID:g556223; PIDN:AAA61174.1; PID:g556224  
R;Kobayashi, K.; Kaneda, N.; Ichinose, H.; Kishi, F.; Nakazawa, A.; Kurosawa, Y.; Fujita  
J. Biochem. 103, 907-912, 1988  
A;Title: Structure of the human tyrosine hydroxylase gene: alternative splicing from a s  
A;Reference number: JE0012; MUID:89008200; PMID:2902075  
A;Accession: JE0012  
A;Molecule type: DNA  
A;Residues: 1-30,62-135 <KOB1>  
A;Cross-references: GB:D00269; NID:g220099; PIDN:BAA25094.1; PID:g2951764  
A;Experimental source: splice form 1  
A;Note: this splice form is produced by an alternative donor site within exon 1  
A;Accession: JE0013  
A;Molecule type: DNA  
A;Residues: 1-34,62-135 <KOB2>  
A;Cross-references: GB:D00269; NID:g220099; PIDN:BAA25097.1; PID:g2951767  
A;Experimental source: splice form 2

A;Accession: JE0014  
A;Molecule type: DNA  
A;Residues: 1-30,35-135 <KOB3>  
A;Cross-references: GB:D00269; NID:g220099; PIDN:BAA25095.1; PID:g2951765  
A;Experimental source: splice form 3  
A;Note: this splice form is produced by an alternative donor site within exon 1  
R;Grima, B.; Lamouroux, A.; Boni, C.; Julien, J.F.; Javoy-Agid, F.; Mallet, J.  
Nature 326, 707-711, 1987  
A;Title: A single human gene encoding multiple tyrosine hydroxylases with different predi  
A;Reference number: A93393; MUID:87173064; PMID:2882428  
A;Accession: A27791  
A;Molecule type: mRNA  
A;Residues: 1-30,62-528 <GRI1>  
A;Cross-references: GB:X05290; NID:g32501; PIDN:CAA28908.1; PID:g32502  
A;Experimental source: splice form 1  
A;Note: this splice form is produced by an alternative donor site within exon 1  
A;Accession: B27791  
A;Molecule type: mRNA  
A;Residues: 1-34,62-528 <GRI2>  
A;Cross-references: GB:X05290; NID:g32501  
A;Experimental source: splice form 2  
A;Accession: C27791  
A;Molecule type: mRNA  
A;Residues: 30,35-528 <GRI3>  
A;Cross-references: GB:X05290; NID:g32501  
A;Experimental source: splice form 3  
A;Note: this isoform is produced by use of an alternative donor site within exon 1  
R;Ichinose, H.; Ohye, T.; Fujita, K.; Yoshida, M.; Ueda, S.; Nagatsu, T.  
Biochem. Biophys. Res. Commun. 195, 158-165, 1993  
A;Title: Increased heterogeneity of tyrosine hydroxylase in humans.  
A;Reference number: P05075; MUID:93371398; PMID:7689834  
A;Accession: P05075  
A;Status: translation not shown  
A;Molecule type: DNA  
A;Residues: 19-30 <CHI>  
A;Accession: P0582  
A;Status: translation not shown  
A;Molecule type: DNA  
A;Residues: 35-61 <CH2>  
A;Accession: P0588  
A;Status: translation not shown  
A;Molecule type: DNA  
A;Residues: 62-106 <ICH3>  
R;O'Malley, K.L.; Anhalt, M.J.; Martin, B.M.; Kelsoe, J.R.; Winfield, S.L.; Ginns, E.I.  
Biochemistry 26, 2910-2914, 1987  
A;Title: Isolation and characterization of the human tyrosine hydroxylase gene: identific  
A;Reference number: I52396; MUID:88107612; PMID:2892528  
A;Accession: I52396  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-61 <OMA>  
A;Cross-references: GB:M18116; NID:g339633; PIDN:AAA77649.1; PID:g1004335  
R;Kobayashi, K.; Kaneda, N.; Ichinose, H.; Kishi, F.; Nakazawa, A.; Kurosawa, Y.; Fujita,  
Neurochem. Res. 15, 6733, 1987  
A;Title: Isolation of a full-length cDNA clone encoding human tyrosine hydroxylase type 1  
A;Reference number: I38340; MUID:87316931; PMID:2888085  
A;Accession: I38340  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-30,35-528 <KOB4>  
A;Cross-references: EMBL:Y00414; NID:g37126; PIDN:CAA68472.1; PID:g37127  
R;Ginns, E.I.; Rehavi, M.; Martin, B.M.; Weller, M.; O'Malley, K.L.; Lamarca, M.E.; McAll  
J. Biol. Chem. 263, 7406-7410, 1988  
A;Title: Expression of human tyrosine hydroxylase cDNA in invertebrate cells using a bac  
A;Reference number: I55282; MUID:88213428; PMID:2896667  
A;Accession: I55282  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-30,62-64 <GIN1>  
A;Cross-references: GB:M20911; NID:g339636; PIDN:AAA61167.1; PID:g339637  
A;Accession: I70056  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA

A:Residues: 1-34,62-64 <GIN2>  
 A:Cross-references: GB:M02912; NID:9339642; PIDN:AAA61168.1; PID:9339643  
 A:Comment: The expression of the four distinct proteins produced by alternate splicing  
 C:Genetics:  
 A:Gene: GDB:TH  
 A:Cross-references: GDB:119612; OMIM:191290  
 A:Map position: 11p15.5,11p15.5  
 A:Introns: 34/3; 61/3; 135/3  
 A:Note: the list of introns is incomplete  
 C:Function:  
 A:Description: catalyzes the 3'-hydroxylation of tyrosine to 3',4'-dihydroxyphenylalanine  
 A:Pathway: catecholamine biosynthesis  
 A:Note: this is the rate-limiting step in catecholamine biosynthesis  
 C:Superfamily: phenylalanine 4-monooxygenase  
 C:Keywords: alternative splicing; biotin; catecholamine biosynthesis; iron; metalloproteins; tyrosine 3-monooxygenase, splice form 4 #status predicted <MAY4>  
 F:1-34,62-528/Product: tyrosine 3-monooxygenase, splice form 4 #status predicted <MAY2>  
 F:1-30,35-528/Product: tyrosine 3-monooxygenase, splice form 3 #status predicted <MAY3>  
 F:1-30,62-528/Product: tyrosine 3-monooxygenase, splice form 1 #status predicted <MAY1>  
 F:6/Binding site: phosphate (Thr) (covalent) (by unidentified kinase) #status predicted  
 F:19/Binding site: phosphate (Ser) (covalent) (by unidentified kinase) #status predicted  
 F:1,183/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #status predicted  
 F:361,366,406/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.6%; Score 220; DB 1; Length 528;  
 Best Local Similarity 26.7%; Pred. No. 6,2e-10;  
 Matches 65; Conservative 38; Mismatches 104; Indels 36; Gaps 5;

QY 112 SRRFSLKMS-----YCRFPDLYEAFGLSDFLDH-----QAVIKFELETHF 155  
 DB 257 AERIATWKEVYTTKGLPATACGHEALPRLRPSGHEBNIPOLEVSSFFLKERTGF 316  
 QY 156 SYYPVSGFVAAPHQVYLSLDQRYFPFASVMTLQDNFSLTPDLIHDLGHVPMILHSPFS 215  
 DB 317 QURPVAGLLSARDPLASLAFRFVFOCTOYIRHASSPMHSPPDCHELLGHVMLADRFPA 376  
 QY 216 EFFIMGRLEFTKVEIKVOALPSKKRIQTLQSNLIAIVRCFWETVESGLIENHGRKAYG 275  
 DB 377 QFSODIG-----LASLQASDEIEKLTLS-----WETVEFGLCCKNGEVKAYG 420  
 QY 276 AVLISSPQELGHAFIDNRYVPLLELDQIRLPFNTSPQETLSIRHF----DELVELTYS 331  
 DB 421 AGLLSYGEHLHCLSEPEIRAPDEPAAVQPDQYGVSVFSESFSDAKDRLSYAS 480  
 QY 332 KLE 334  
 DB 481 RTQ 483

RESULT 16  
 A:Residues: 1-579  
 A:Cross-references: EC 1.14.16.2, major splice form - fruit fly (Drosophila melanogaster)  
 N:Contains: tyrosine 3-monooxygenase, type II; tyrosine 3-hydroxylase, type II  
 C:Species: Drosophila melanogaster  
 C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text\_change 21-Jul-2000  
 C:Accession: A55369; B55369; JN0010  
 R:Bitman, S.; Morgan, B.; Anzivino, M.; Hirsch, J.  
 J. Biol. Chem. 269, 26559-26567, 1994  
 A:Title: A novel and major isoform of tyrosine hydroxylase in Drosophila is generated by  
 A:Reference number: A55369; MUID:95014502; PMID:7929381  
 A:Accession: A55369  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-579 <BR1>  
 A:Cross-references: GB:U14395; NID:9595799; PIDN:AAA62876.1; PID:9595800  
 A:Note: authors translated the codon GCG for residue 219 as Ser, and TCC for residue 220  
 A:Accession: B55369  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-61,133-579 <BR2>  
 A:Cross-references: GB:U14395; NID:9595799; PIDN:AAA62877.1; PID:9595801  
 A:Note: authors translated the codon GCG for residue 219 as Ser, and TCC for residue 220

R:Neckameyer, W.S.; Quim, W.G.  
 Neuron 2, 1167-1175, 1989  
 A:Title: Isolation and characterization of the gene for drosophila tyrosine hydroxylase  
 A:Reference number: JN0010; MUID:90166583; PMID:2483109  
 A:Accession: JN0010  
 A:Molecule type: mRNA  
 A:Residues: 1-61,133-579 <NEC>  
 A:Cross-references: GB:X76209; NID:9433469; PIDN:CAA53802.1; PID:9433470  
 C:Genetics:  
 A:Gene: FlyBase:ple  
 A:Cross-references: FlyBase:FBgn0005626  
 A:Map position: 3L 65B  
 A:Introns: 23/3; 61/1; 101/1; 132/1; 405/2; 447/3  
 C:Superfamily: phenylalanine 4-monooxygenase  
 C:Keywords: alternative splicing; biotin; iron; metalloprotein; oxidoreductase  
 F:409,414,454/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.6%; Score 219; DB 1; Length 579;  
 Best Local Similarity 28.6%; Pred. No. 8.4e-10;  
 Matches 74; Conservative 37; Mismatches 94; Indels 54; Gaps 8;

QY 128 DYLEAFGLSD----FLDH-----QAVIKFELETHFSSYYPVSGFVAAPHQVYLSLDQRYFP 179  
 DB 329 EYRAAFQKQDBQIFVETRLPQLQEMSDFLKRTGFGSLRPAAGLTPARPLASLAFRIQ 388  
 QY 180 IASVMTLQDNFSLTPDLIHDLGHVPMILHSPFSFPIFMGRLPFKYIEKVOALPSKK 239  
 DB 389 STQYRVAVNPPYTPPDSIHDLGHVPMILHSPFSFQSGEIG-----LASLQASD 439  
 QY 240 QRIOTLQSNLIAIVRCFWETVESGLIENHGRKAYGAVLISPOELGHAFID----- 291  
 DB 440 EIEIKST-----VWFTVEFGLCCKEGRQIKAYGGLSSYGEHLHAISSDKCEHRAFE 492  
 QY 292 --NVRLVPLE-----LDQIRLPFNTSPQETLSIRHF--ELVELT 330  
 DB 493 PASTAVQPDQYGVSVFSESFSDAKDRLSYAS  
 QY 331 SKLEWMLDQGLLESIPLYN 349  
 DB 552 DKLETLVQKMTLEIHLTN 570

RESULT 17  
 A:Residues: 1-453 <DAB>  
 A:Cross-references: GB:M12337; NID:9206120; PIDN:AAA41843.1; PID:9206121  
 R:Robson, K.J.H.; Battie, W.; James, R.J.; Cotton, R.C.H.; Morgan, F.J.; Woo, S.L.C  
 Biochemistry 23, 5671-5675, 1984  
 A:Title: Sequence comparison of rat liver phenylalanine hydroxylase and its cDNA clo  
 A:Reference number: A00509; MUID:85122617; PMID:6098294  
 A:Accession: A00509  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 208-453 <ROB>  
 A:Cross-references: GB:K02599; NID:9205961; PIDN:AAA41794.1; PID:9205962  
 R:Woborn, M.; Humble, E.; Ragnarsdottir, U.; Engstrom, L.  
 Biochem Biophys Res Commun 93, 403-408, 1980  
 A:Title: Amino acid sequence at the phosphorylated site of rat liver phenylalanine h  
 A:Reference number: A14970; MUID:80220293; PMID:7387651  
 A:Accession: A14970  
 A:Molecule type: protein  
 A:Residues: 12-16, 'B', '18-19', 'ZZ' <WRB>  
 C:Complex: homodimer



C:Function:  
A:Description: catalyzes the 4'-hydroxylation of phenylalanine to tyrosine by tetrahydrofuran  
A:Pathway: tyrosine biosynthesis; phenylalanine catabolism  
C:Superfamily: phenylalanine 4-monooxygenase  
C:Keywords: bioterror; homodimer; iron; metalloprotein; monooxygenase; oxidoreductase; p  
F:16/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #status p  
F:23/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status predict  
F:285,290,330/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.3%; Score 214; DB 1; Length 453;  
Best Local Similarity 28.4%; Pred. No. 1.5e-09;  
Matches 61; Conservative 37; Mismatches 89; Indels 28; Gaps 4;

QY 115 FSLMKSVCPRFLDYLEAFGLLSDFLDH-OAVIKFELETHFSYYPVSGFVAPHQYLSLL 173  
DB 210 FPLEKYC-----CFREDNPQLEDVVSQFLQCTGFRLRPVAGLLSRDPLGL 258

QY 174 QDRYPTASVMRTLDKNFSLTDLHLLGHVPWLLHPSFSEFFINMGRLFTKVIKQV 233  
DB 259 APRVPHCTQYIRHSGKPMYTPEDICHLLGHVPLFSDRSPAQFSQIEG-----LA 309

QY 234 ALPSKKQRIOTLQSNLIAIVRCFWFTVESGLIENHGRKAYGAVLSSPOLGHAFIDNV 293  
DB 310 SLGAPDEVIEKLTAT-----IYWFTEVFGLCCKGDSIKAYGAGLLSSFGELQYCLSDKP 362

QY 294 RVLPLELDQIIRLPENTSTPOETLFSIRHFDLVE 328  
DB 363 KLLPLELEKTAQCEYSVTETQPLYYVAESFSDAKE 397

RESULT 18  
WHUF  
phenylalanine 4-monooxygenase (EC 1.14.14.16.1) - human  
N:Alternate names: phenylalanine 4-hydroxylase  
C:Species: Homo sapiens (man)  
C>Date: 28-Aug-1985 #sequence\_revision 28-Aug-1985 #text\_change 21-Jul-2000  
C:Accession: A00508; S02687; I52416; I54346; S74142; I54257  
R:Kwok, S.C.M.; Ledley, F.D.; Dilella, A.G.; Robson, K.J.H.; Woo, S.L.C.  
Biochemistry 24, 556-561, 1985  
A:Title: Nucleotide sequence of a full-length complementary DNA clone and amino acid seq  
A:Reference number: A00508; MUID:85199778; PMID:2986678  
A:Accession: A00508  
A:Molecule type: mRNA  
A:Residues: 1-452 <KWO>  
A:Cross-references: GB:K03020; NID:G189936; PIDN:AAA60082.1; PID:G189937  
R:Cotton, R.G.H.; McAdam, W.; Jennings, I.; Morgan, F.J.  
Biochem. J. 255, 193-196, 1988  
A:Title: A monoclonal antibody to aromatic amino acid hydroxylases. Identification of th  
A:Reference number: S02687; MUID:89061656; PMID:2461704  
A:Accession: S02687  
A:Molecule type: protein  
A:Residues: 131-144 <COF>  
R:Konecki, D.S.; Wang, Y.; Trefz, F.K.; Lichter-Konecki, U.; Woo, S.L.  
Biochemistry 31, 8363-8368, 1992  
A:Title: Structural characterization of the 5' regions of the human phenylalanine hydrox  
A:Reference number: I52416; MUID:92399453; PMID:1326329  
A:Accession: I52416  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-20 <KON>  
A:Cross-references: GB:S44225; NID:G255493  
R:Abadie, V.; Jaruzelska, J.; Lyonnet, S.; Millasseau, P.; Berthelon, M.; Rey, F.; Munni  
Hum. Mol. Genet. 2, 31-34, 1993  
A:Title: Illegitimate transcription of the phenylalanine hydroxylase gene in lymphocytes  
A:Reference number: I54346; MUID:93258345; PMID:8098245  
A:Accession: I54346  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 381-389, 'C' 391-405 <ABR>  
A:Cross-references: GB:S61296; NID:G300410; PIDN:AAD13926.1; PID:94261626  
R:Kowlessur, D.; Citron, B.A.; Kaufman, S.  
Arch. Biochem. Biophys. 333, 85-95, 1996

A:Title: Recombinant human phenylalanine hydroxylase: novel regulatory and structural pr  
A:Reference number: S74142; MUID:96400381; PMID:8806757  
A:Accession: S74142  
A:Molecule type: protein  
A:Residues: 2-21 <KOW>  
R:Eigel, A.; Dworniczak, B.; Kalaydjieva, L.; Horst, J.  
Hum. Genet. 87, 739-744, 1991  
A:Title: A frameshift mutation in exon 2 of the phenylalanine hydroxylase gene linked to  
A:Reference number: I54257; MUID:92039642; PMID:1682235  
A:Accession: I54257  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 53-54, 'LRRMM' <EIG>  
A:Cross-references: GB:S62592; NID:G238240; PIDN:AAB20205.1; PID:G238241  
A>Note: a defect in this gene can cause phenylketonuria  
C:Genetics:  
A:Gene: GDB:PAH  
A:Cross-references: GDB:119470; OMIM:261600  
A:Map position: 12q24.1-12q24.1  
A>Note: a defect in this gene can cause phenylketonuria  
C:Complex: homodimer  
C:Function:  
A:Description: catalyzes the 4'-hydroxylation of phenylalanine to tyrosine by tetrahydrof  
A:Pathway: tyrosine biosynthesis; phenylalanine catabolism  
C:Superfamily: phenylalanine 4-monooxygenase  
C:Keywords: bioterror; homodimer; iron; metalloprotein; monooxygenase; oxidoreductase; p  
F:16/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #status p  
F:23/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status predict  
F:285,290,330/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.3%; Score 213.5; DB 1; Length 452;  
Best Local Similarity 26.6%; Pred. No. 1.7e-09;  
Matches 62; Conservative 44; Mismatches 100; Indels 27; Gaps 4;

QY 104 RNLMYRLSSGRFSLWKSVCPRFLDYLEAFGLLSDFLDH-----OAVIKFELETHP 155  
DB 184 KKTGTGTVFKLTKSLYKTHA---CYENHIFPLEKYCGFHEDNPQLEDVVSQFLQCTGF 240

QY 156 SYVPVSGFVAPHQYLSLDQRYPTASVMRTLDKNFSLTDLHLLGHVPWLLHPSFS 215  
DB 241 RLRPVAGLLSRDPLGLAFRVPHCTQYIRHSGKPMYTPEDICHLLGHVPLFSDRSPA 300

QY 216 EPPINMGRLFTKVIKQVQALPSKKQRIOTLQSNLIAIVRCFWFTVESGLIENHGRKAYG 275  
DB 301 QFSQIEG-----LASLGAPDEVIEKLTAT-----IYWFTEVFGLCCKGDSIKAYG 344

QY 276 AVLSSPOLGHAFIDNVRLPDLQIIRLPENTSTPOETLFSIRHFDLVE 328  
DB 345 AGLSSFGELQYCLSEKPKLLPLELEKTAQYVTETQPLYYVAESFSDAKE 397

RESULT 19  
S15758  
phenylalanine 4-monooxygenase (EC 1.14.14.16.1) - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 31-Mar-2000  
C:Accession: S15758; S08678  
R:Ledley, F.D.; Grenett, H.E.; Dunbar, B.S.; Woo, S.L.C.  
Biochem. J. 267, 399-406, 1990  
A:Title: Mouse phenylalanine hydroxylase. Homology and divergence from human phenylalanin  
A:Reference number: S15758; MUID:90241147; PMID:2334400  
A:Accession: S15758  
A:Molecule type: mRNA  
A:Residues: 1-453 <LED>  
A:Cross-references: EMBL:X51942; NID:G53683; PIDN:CAA36205.1; PID:G53684  
C:Superfamily: phenylalanine 4-monooxygenase  
C:Keywords: bioterror; iron; metalloprotein; oxidoreductase; phosphoprotein  
F:285,290,330/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.2%; Score 212.5; DB 2; Length 453;  
Best Local Similarity 29.9%; Pred. No. 2e-09;  
Matches 61; Conservative 35; Mismatches 87; Indels 21; Gaps 5;

QY 104 RNLM---YRLSSRFSLMKSCPRFFLDYLEAFGLSDPLDH---QAVIKFELETHFSY 158  
DB 184 RKTWGTFRRLKAKYKTHACYEHHNIFPLLEKICGFRDNIPQLEDSVDFLOTCTGFRLR 243  
QY 159 PVSGFVAHQYLSLQDRYFPFIASVMTLKDKNFSLTPDLIDHGLGVPMILHPSFSEFF 218  
DB 244 PVAGLLSRDFLGLAFRRVFCQYIRHSGKPMWTEPDPICHELLGHVPLPSDRSFAQFS 303  
QY 219 IMMGRLEFTKVIKQVALPESKKORIQTQSNLAIIVRCFWFTVESGLIENHGRKAYGAVL 278  
DB 304 QEIG-----LASLGAPDEYIEKLTAT-----IWFVTFEGICKEDSDIAYAGAGL 347  
QY 279 ISSPOLGHAFLDNVRLVPLELQD 302  
DB 348 ISSFGELOGLSDKPKLLPLELEK 371

## RESULT 20

tyrosine 3-monooxygenase (EC 1.14.16.2) - bovine  
N/Alternate names: tyrosine 3-hydroxylase  
C/Species: Bos primigenius laurus (cattle)  
C/Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 31-Mar-2000  
A/Accession: J00039; A27687  
R/Motals: S.; Stenle, A.D.; Lamouroux, A.; Mallet, J.; Thoenen, H.  
J. Neurochem. 51, 572-578, 1988  
A/Title: Predicted amino acid sequence of bovine tyrosine hydroxylase and its similarity  
A/Reference number: J00039; MUID:88274405; PMID:2899135  
A/Accession: J00039  
A/Molecule type: mRNA  
A/Residues: 1-491 <SAA>  
A/Cross-references: GB:M36705; NID:G163796; PIDN:AAA0798.1; PID:G163797  
A/Experimental source: adrenal medulla  
R/Abate, C.; Smith, J.A.; Joh, T.H.  
Biochem. Biophys. Res. Commun. 151, 1446-1453, 1988  
A/Title: Characterization of the catalytic domain of bovine adrenal tyrosine hydroxylase  
A/Reference number: A27687; MUID:88183482; PMID:2895648  
A/Accession: A27687  
A/Molecule type: protein  
A/Residues: 153-158,160,162-169 <ABA>  
C/Comment: This protein is the rate-limiting enzyme in the biosynthesis of catecholamine  
C/Superfamily: phenylalanine 4-monooxygenase  
C/Keywords: bioprotein; iron; metalloprotein; oxidoreductase; phosphoprotein  
F/2-491/Product: tyrosine 3-monooxygenase #status predicted <MA>  
F/8/Binding site: phosphate (Ser) (covalent) (by unidentified kinase) #status predicted  
F/19/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #status P  
F/40/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status predict  
F/324,329,369/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.2%; Score 211.5; DB 2; Length 491;

Best Local Similarity 26.7%; Pred. No. 2.7e-09;

Matches 64; Conservative 41; Mismatches 104; Indels 31; Gaps 5;

QY 107 WYRLSSRFSLMKSCPRFFLDYLEAFGLSDPLDH-----QAVIKFELETHFSY 158  
DB 126 WKEXVSTRGLYPTTHAC---EHLAELELLERFCGYREDRIPQLEDSVDFLOTCTGFR 282  
QY 159 PVSGFVAHQYLSLQDRYFPFIASVMTLKDKNFSLTPDLIDHGLGVPMILHPSFSEFF 218  
DB 283 PVAGLLSRDFLGLAFRRVFCQYIRHSGKPMWTEPDPICHELLGHVPLPSDRSFAQFS 342  
QY 219 IMMGRLEFTKVIKQVALPESKKORIQTQSNLAIIVRCFWFTVESGLIENHGRKAYGAVL 278  
DB 343 QDIG-----LASLGVSDDELEKLT-----LYMFVTFEGICKEDSDIAYAGAGL 386  
QY 279 ISSPOLGHAFLDNVRLVPLELQDIIIRLPNTSTPOETLPSIRH---DELVELTSGLE 334  
DB 387 ISSYGELEHLSSEBEIRAFDPDAAVQPDQOTQYVYFVSESPDAXDKLRYSYKIQ 446

## RESULT 21

J00766  
phenylalanine 4-monooxygenase (EC 1.14.16.1) - fruit fly (Drosophila melanogaster)

N/Alternate names: phenylalanine 4-hydroxylase; tryptophan hydroxylase  
C/Species: Drosophila melanogaster  
C/Date: 10-Mar-1994 #sequence\_revision 10-Mar-1994 #text\_change 31-Mar-2000  
A/Accession: J00766  
R/Motals: G.; Requena, J.M.; Jimenez-Ruiz, A.; Lopez, M.C.; Ugarte, M.; Alonso, C.  
Gene 93, 213-219, 1990  
A/Title: Sequence and expression of the Drosophila phenylalanine hydroxylase mRNA.  
A/Reference number: J00766; MUID:91033030; PMID:2121612  
A/Accession: J00766  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-453 <MOR>  
A/Cross-references: GB:M32802; NID:G158035; PIDN:AA69513.1; PID:G158036  
C/Genetics:  
A/Gene: FlyBase:TpH  
A/Cross-references: FlyBase:FBgn0005770  
C/Superfamily: phenylalanine 4-monooxygenase  
C/Keywords: bioprotein; iron; metalloprotein; monooxygenase; oxidoreductase; phenylke  
F/285,290,330/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.0%; Score 207; DB 2; Length 453;

Best Local Similarity 27.9%; Pred. No. 5.7e-09;

Matches 75; Conservative 39; Mismatches 109; Indels 46; Gaps 10;

QY 107 WYRLSSRFSLMKSCPRFFLDYLEAFGLSD---FLDH-----QAVIKFELETHFSY 158  
DB 187 WGIIFRNLYTKLYTHACR---SYNHVPLVDNCGFRDNIPQLEDSVDFLOTCTGFR 243  
QY 159 PVSGFVAHQYLSLQDRYFPFIASVMTLKDKNFSLTPDLIDHGLGVPMILHPSFSEFF 218  
DB 244 PVAGLLSRDFLGLAFRRVFCQYIRHSGKPMWTEPDPICHELLGHVPLPSDRSFAQFS 303  
QY 219 IMMGRLEFTKVIKQVALPESKKORIQTQSNLAIIVRCFWFTVESGLIENHGRKAYGAVL 278  
DB 304 QEIG-----LASLGAPDDYIEKLT-----IWFVTFEGILAKGELKAYAGAGL 347  
QY 279 ISSPOLGHAFLDNVRLVPLELQDIIIRLPNTSTPOETLPSIRHPELVETSKLEMD 338  
DB 348 ISSYGELEHLSSEBEIRAFDPDAAVQPDQOTQYVYFVSESPDAXDKLRYSYKIQ 397  
QY 339 QGV---LESTP-----LYNDEKLSGFEVL 360  
DB 398 KTIKFANSIRPRGVKYN---ATQSVLEV 424

## RESULT 22

hypothetical protein ZK1290.2 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
A/Accession: T34509  
submitted to the EMBL Data Library, July 1995  
A/Description: The sequence of C. elegans cosmid ZK1290.  
A/Reference number: Z21535  
A/Accession: T34509  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-575 <TAI>  
A/Cross-references: EMBL:U21308; PIDN:AA93319.1; GSPDB:GN00020; CESP:ZK1290.2  
A/Experimental source: strain Bristol N2; clone ZK1290  
C/Genetics:  
A/Map position: 2  
A/Introns: 49/3; 72/2; 105/3; 223/1; 257/3; 280/2; 310/2; 346/1; 391/2; 433/3; 523/2

Query Match 10.9%; Score 206.5; DB 2; Length 575;

Best Local Similarity 25.2%; Pred. No. 8.5e-09;

Matches 70; Conservative 47; Mismatches 116; Indels 45; Gaps 8;

QY 104 RNLM---YRLSSRFSLMKSCPRFFLDYLEAFGLSDPLDH-----QAVIKFELETHFSY 155  
DB 294 RKTWGIITRRLK---ELHKKHACKOPFLDNFELRHGCGYSBNIPQLEDCIKFLKAKTGF 350



QY 215 SFFINMRLFTKVIKVOALPSKQRIQTLQSNLAIYRCFMTVESGLIENHGKAY 274  
DB 144 ADY-----MAYGEGGRALGL-GRNLNLRKLYWYVERGLNMTFAGLRIRY 188  
QY 275 GAVLISSPOELGHAFFID-NRVVLEPLEDQIIRLPFNTSTPQSTLSIHGPELVLT--- 330  
DB 189 GAGIVSSKRTSIFALDDPSRIGFDLERVMTLYRIDFQOYFVIDSIQTLQDVTLRD 248  
QY 331 -----SKLEWMLDQGLLESIP 346  
DB 249 FGAIVYERLASVSDIGVAEIVP 269

## RESULT 26

T25453  
tyrosine 3-monooxygenase (EC 1.14.16.2) B0432.5 [similarity] - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 19-May-2000  
C/Accession: T25453  
R/Henhaus, J.; Wohldmann, P.  
submitted to the EMBL Data Library, December 1996  
A/Description: The sequence of C. elegans cosmid B0432.  
A/Reference number: Z20038  
A/Accession: T25453  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-404 <HEN>  
A/Cross-references: EMBL:U80836; PIDN:AA837888.1; GSPDB:GN00020; CESP:B0432.5  
A/Experimental source: strain Bristol NZ; clone B0432  
C/Genetics:  
A/Genes: CESP:B0432.5  
A/Map position: 2  
A/Introns: 117/2; 183/1; 273/2; 302/3; 369/3  
C/Superfamily: phenylalanine 4-monooxygenase  
C/Keywords: bioplerin; oxidoreductase

## Query Match

Best Local Similarity 24.4%; Score 178.5; DB 2; Length 404;  
Matches 77; Conservative 41; Mismatches 114; Indels 85; Gaps 11;

QY 19 KLRQSLSPFQNSQSLORAVSTPYSYRIILQKENKQA-LARH-----KCTISILE 69  
DB 32 QLIHSAELTLQNHVAL-----TKSFIAKLSDEKQSGIWPFRHISLQCSKCIKRYE 86  
QY 70 FPKNLFPVHLSTLKNQREGSTDMVAVSTPPFNRLMWRLSSRPSLMKSCYCPFLDY 129  
DB 87 -----PTDRPHGHQHDVAVIARKTLND--QALEFRFGDEIGY-----VDY 126  
QY 130 LE-----AFGLSDP-LDHOAVI-----KPELETH 154  
DB 127 TEEHATWKAIVTEKLDLHSHTCAYRQNLKILQSEKVLTDRIIPQIRDVNKKFLQCKTG 186  
QY 155 PSYVPSGFAVPHQYLSLQDRYPPIASVMTLDKNFSLTPDLIHLLGHVPMWLHPSP 214  
DB 187 FELRCGSLSLARDFLASLAFRVQTTTYLRHKKSPHNSPEPDLIHLLGHVPMFSDPL 246  
QY 215 SEFLINMGRFLTQVIEKVOALPSKQRIQTLQSNLAIYRCFMTVESGLIENHGKAY 274  
DB 247 AQMSQDIG-----LMSLGASDEHIKLSF-----VVMFIVERGLCKEDGKAKAI 290  
QY 275 GAVLISSPOELGHAFFID 291  
DB 291 GAGLIASVGEIMHACSD 307

## RESULT 27

T23494  
phenylalanine 4-monooxygenase (EC 1.14.16.1) K0898.4 [similarity] - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1989 #sequence\_revision 15-Oct-1999 #text\_change 19-May-2000  
C/Accession: T23494  
R/Smey, R.  
submitted to the EMBL Data Library, October 1995

A/Reference number: Z19747  
A/Accession: T23494  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-457 <WIL>  
A/Cross-references: EMBL:Z66497; PIDN:CAA91286.1; GSPDB:GN00020; CESP:K0898.4  
A/Experimental source: clone K0898  
C/Genetics:  
A/Genes: CESP:K0898.4  
A/Map position: 2  
A/Introns: 12/3; 51/3; 116/1; 170/2; 236/1; 281/2; 323/3; 400/2  
C/Superfamily: phenylalanine 4-monooxygenase  
C/Keywords: bioplerin; oxidoreductase

## Query Match

Best Local Similarity 25.5%; Score 175.5; DB 2; Length 457;  
Matches 56; Conservative 35; Mismatches 80; Indels 49; Gaps 5;

QY 143 QAVIKPELETHPSYVPSGFAVPHQYLSLQDRYPPIASVMTLDKNFSLTPDLIDY 202  
DB 228 QDVSDFLKQCTGYTIRPVAAGLSSRDPLAGLAFVHSTGYIRHSAKPYTPEPDI 287  
QY 203 LGHVPMWLHPSPSEFFINMGRF-----TKVIEKVOALPSKQRIQTLQSNLAIYRCMF 258  
DB 288 LGHVPLFADVERFQPSQETGLASGAPDVIETKATL-----YVF 327  
QY 259 TVESGLIENHGKAYGAVLISSPOELGHAFFIDNRVLE--LDQIIRLPFNTSTPQST 316  
DB 328 TIERGICQDGEKKAAGLSSFGELOYALSDPEVDDPAVCCVTXYPITTEYQPKYF 387  
QY 317 L-----FSIRH--FDELVELTSKL 333  
DB 388 LAESFASANKKLSMAATINRPFQIRYNAVQREVLIDKY 427

## RESULT 28

T45787  
disease resistance-like protein - Arabidopsis thaliana  
N/Alternate names: protein F26013.200  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
R/Delsany, M.; Berger, C.; Cooke, R.; Grellett, F.; Laurie, M.; Mewes, H.W.; Lemcke, F.  
submitted to the Protein Sequence Database, December 1999  
A/Reference number: Z23013  
A/Accession: T45787  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-1253 <DEU>  
A/Cross-references: EMBL:AL133452  
A/Experimental source: cultivar Columbia; BAC clone F26013  
C/Genetics:  
A/Map position: 3  
A/Introns: 138/2; 490/3; 589/3; 1047/3  
A/Note: F26013.200

## Query Match

Best Local Similarity 5.8%; Score 109.5; DB 2; Length 1253;  
Matches 81; Conservative 61; Mismatches 130; Indels 127; Gaps 18;

QY 77 VHLSTLSTKNQREGS-----TDMAVSTPFFNRNLMYLLSS-----RF 115  
DB 421 IHL-----EGCGFFPVEINVLVEKCLVSNM-EGRVVMHNLIGSIRKTIINGKRRS 472  
QY 116 SLMKSCYCPFFLDYLEARG-----LSDPFLDHOAV-----IKFFEL-----ET 153  
DB 473 RUKPKLIRKTFLEDRQVGSDEIAIFLDPBALSDVDNPAFEMNTNNAKYLKICSSNGN 532  
QY 154 HPSYVPSGFAVPHQYLSLQDRYPPIASVMTLDKNFSL-----TPD-- 197  
DB 533 HYALHLPKGVKSPEBLRLHWEHFPDLISLQDQNTNNTNVLIMCKYSKQLRMBGTEKG 592  
QY 198 -LTHDLLGHVPMWLHPSPSEFFINM-----GRL-----FTVYIE-----KVQ 233

Db 593 MLKRLMLCHSQQLVIGIQELQIALNMEVIDLQCARLQRLATGHQHLRVINLSGCIK 652  
QY 234 ALPSKKKRIOTL---QSNLAIIVRCFWFVTEGLLENHGRKAYGAVLSSPOELG-HAF 289  
Db 653 SPFEPVPIEELKQGIKIRISPTVTFSPQDINSFYDHDKHFLNREVSSDSQSLSIMVY 712  
QY 290 IDNVRLP---LELDQIRLPNT-----STPQTLFSTRHDELVEL---TSKLEWM 336  
Db 713 LDNLKVLDSQLELDIQIPKRLKYLGTAKELPSLMHLSLVLDLENCKRLHK 772  
QY 337 LDQGL-----LESIPLYNQKYLSSG 356  
Db 773 LPMGIGNLSSVLNLSGCGSELDIQIPRNLEELYLAG 811  
RESULT 29  
B29813  
174K ninaC protein - fruit fly (*Drosophila melanogaster*)  
N:Contains: protein kinase (EC 2.7.1.-)  
C:Species: *Drosophila melanogaster*  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 19-Jan-2001  
C:Accession: B29813  
R:Montell, C.; Rubin, G.M.  
Cell 52, 757-772, 1988  
A:Title: The *Drosophila* ninaC locus encodes two photoreceptor cell specific proteins with  
A:Reference number: A90898; MUID:88151067; PMID:2449973  
A:Accession: B29813  
A:Molecule type: mRNA  
A:Residues: 1-1501 <MON>  
A:Cross-references: GB:M20230; NID:q157967; PIDN:AAA28721.1; PID:q157968  
C:Genetics  
A:Gene: FlyBase:ninaC  
A:Cross-references: FlyBase:FBgn0002938  
C:Superfamily: ninaC protein; myosin motor domain homology; protein kinase homology  
C:Keywords: actin binding; alternative splicing; ATP; nucleotide binding; P-loop; phosph  
F:14-282/Domain: protein kinase homology <kin>  
F:335-1022/Domain: myosin motor domain homology <MMOT>  
F:425-432/Region: nucleotide-binding motif A (P-loop)  
F:911-936/Region: actin binding #status predicted  
F:1054-1501/Domain: carboxyl-terminal <CBT>  
F:45,60,145/Active site: Lys, Glu, Asp #status predicted  
F:431/Binding site: ATP (Lys) #status predicted  
Query Match 5.7%; Score 108.5; DB 1; Length 1501;  
Best Local Similarity 22.4%; Pred. No. 2.4;  
Matches 85; Conservative 58; Mismatches 130; Indels 107; Gaps 21;  
QY 10 PKVILKIALKRLSFLFFONSQSLQRAYSTPVSYYRILOKENKEKOALRHK---CI 65  
Db 839 PEMIETFRSSLDSEIMLMFTNLTQKAGNLTP---FEAVQHKDSERKSYALNTLSAGCI 895  
QY 66 SILEFFKLLFVHLISLKNORCGCTDMVYSTPFFNRLWYRLSS-RFSLWKSVCPR 124  
Db 896 SQVN-----NURLANFRCTLTLMKLSQ---NANLOVHFVRCIRADL--EYKPR 942  
QY 125 FF-----LDYLEAFGLISDFLDHQAVIKFPELETHSYYPVSGFVAPHQYLSLLQDRYFP 179  
Db 943 SFHSDVQOQMKALGVLDTVIAKQ---KGS-----SRLPDFELRYQFLAFDFDE--P 992  
QY 180 IASVMTLDKDNFSLTDPDLIDL-----IGHVPMLLHPSSEFFINNGRLFTKVIKQV 233  
Db 993 V-----EMTKNCRSL---LFRLKMGESWALGKTGKFLRYYNDEF---LRLYLELOVKVY- 1040  
QY 234 ALPSKKKRIQTLQSNLAIIVRCFWFVTEGLLENHGRKAYGAVLSSPOEL 285  
Db 1041 -----IKVQSMRALLARKR-----VKGGKVPKLGKGPEDHD-----VAASKIQKA 1082  
QY 286 GHAFIDNVRLPLELDQIRL-----PNTSTPQTLFSI-----RHFDDELVELT 330  
Db 1083 FRGFRDRLVPLVNEKSGQLNENTADFIKPAKKWREKSIQVLLHYRAARQDFNLIS 1142  
QY 331 SKLEWMLDQGLLESIPLYNQ 350

Db 1143 -----QQVHIYNQ 1150  
RESULT 30  
A40996  
phenylalanine 4-monooxygenase (EC 1.14.16.1) - *Chromobacterium violaceum*  
N:Alternate names: phenylalaninase; phenylalanine 4-hydroxylase  
C:Species: *Chromobacterium violaceum*  
C:Date: 03-Aug-1992 #sequence\_revision 03-Aug-1992 #text\_change 31-Mar-2000  
C:Accession: A40996; B40996  
R:Onishi, A.; Liotta, L.J.; Benkovic, S.J.  
J. Biol. Chem. 266, 18454-18459, 1991  
A:Title: Cloning and expression of *Chromobacterium violaceum* phenylalanine hydroxylase in  
A:Reference number: A40996; MUID:92011593; PMID:1655752  
A:Accession: A40996  
A:Molecule type: DNA  
A:Residues: 1-296 <ONI>  
A:Cross-references: GB:M55915; NID:g144481; PIDN:AAA23115.1; PID:g144482  
A:Accession: B40996  
A:Molecule type: protein  
A:Residues: 1-20 <ONS>  
A:Experimental source: strain ATCC 12540  
C:Keywords: biotpterin; oxidoreductase  
Query Match 5.6%; Score 106.5; DB 2; Length 296;  
Best Local Similarity 27.3%; Pred. No. 0.4;  
Matches 27; Conservative 22; Mismatches 49; Indels 1; Gaps 1;  
QY 126 FLDYLEAFGLISDFL-DHQAVIKFPELETHSYYPVSGFVAPHQYLSLLQDRYFPPIASVM 184  
Db 63 FLEGLERLEVDADRDVDFNKLNEKLMKMAATKIVAVPGLIPDDVFPEHLANRRFPVTWWL 122  
QY 185 RTLDKDNFSLTDPDLIDLGHVPMLLHPSSEFFINNGR 223  
Db 123 REPQLDLYQEPDVFHDLFGHVLPLINPVFADILEAYGK 161  
RESULT 31  
S44054  
genome polyprotein - Marburg virus (strain Popp)  
N:Alternate names: structural protein L  
N:Contains: RNA-directed RNA polymerase (EC 2.7.7.48)  
C:Species: Marburg virus  
A:Variety: strain Popp  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 18-Jun-1999  
C:Accession: S44054; S32776  
R:Bukreyev, A.A.; Volchkov, V.E.; Blinov, V.M.; Dryga, S.A.; Netesov, S.V.  
submitted to the EMBL Data Library, January 1994  
A:Description: Full-length nucleotide sequence of Marburg virus Popp strain: The comparis  
A:Reference number: S44049  
A:Accession: S44054  
A:Molecule type: genomic RNA  
A:Residues: 1-2331 <BUK>  
A:Cross-references: EMBL:Z29337; NID:g450908; PIDN:CAA82542.1; PID:g450915  
A:Experimental source: strain Popp  
R:Bukreyev, A.A.; Netesov, S.V.  
submitted to the EMBL Data Library, September 1992  
A:Description: The partial nucleotide sequence of Marburg virus genome.  
A:Reference number: S32775  
A:Accession: S32776  
A:Molecule type: genomic RNA  
A:Residues: 1-2331 <BUW>  
A:Cross-references: EMBL:X68494; NID:g296962; PIDN:CAA48508.1; PID:g296963  
A:Experimental source: strain Popp  
C:Genetics  
A:Gene: L  
C:Superfamily: parainfluenza virus RNA-directed RNA polymerase  
C:Keywords: ATP; nucleotidyltransferase  
Query Match 5.5%; Score 104.5; DB 2; Length 2331;  
Best Local Similarity 23.3%; Pred. No. 9.1;  
Matches 59; Conservative 42; Mismatches 117; Indels 35; Gaps 7;

QY 11 KYIKALKLNQSLSLFQNSQSLQRAVSTPVSYYRIILQKENEKQALAHKICISILF 70  
 DB 356 KTIIOGKLTPOQYCELF-----SLQKMGHPVLIVDIALDKYKAQASVKILKPKWPFET 410  
 QY 71 FQNLFFVHLISLKNQREGCSTDMAVAVSTPFNNRLMYRLSSRFSLMKSCPRFLDYTL 130  
 DB 411 FQVFKFIYAKOHVHSSQSWYKTMTHLTPYLRQHIVSNSPSSQAEIYQOHLMEWYFYVHE 470  
 QY 131 EAFGL-----LSDFLDHQAIVKFELETHFSYVPVGFAPHOVYSLQDRYFPIASVNR 185  
 DB 471 PLFSTKIISLDSIFIKORATVANOECWDSVFRSIVGYNPVRFGSKVPEPF----- 523  
 QY 186 TIDKDNFSLTPDLIHDHLPMLHSPSEPF-----INMGELFKVIEKQALPSK 238  
 DB 524 -LQADPSL--NQLIDFMEKLEIYA-PSYRNFSPSLKEKELNIGRTGK-----LPIR 572  
 QY 239 KQRIOTLOSNIYA 251  
 DB 573 VRRVQTLAVALIA 585

## RESULT 32

T38766  
 probable transcription regulatory protein - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 02-Aug-2002  
 C:Accession: T38766  
 R:Gentles, S.; Churcher, C.M.; Barrall, B.G.; Rajandream, M.A.; Walsh, S.V.  
 submitted to the EMBL Data Library, January 1996  
 A:Reference number: Z21810  
 A:Accession: T38766  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-563 <GEN>  
 A:Cross-references: EMBL:Z69086; PIDN:CAA93165.1; GSPDB:GN00066; SPDB:SPAC3H8.08C  
 C:Experimental source: strain 972h-; cosmid c3H8  
 C:Genetics:  
 A:Gene: SPDB:SPAC3H8.08C  
 A:Map position: 1  
 C:Superfamily: transcription regulator SPEC16G5.17; GAL4 zinc binuclear cluster homology  
 F:14-50/Domain: GAL4 zinc binuclear cluster homology <GAL4>

Query Match 5.4%; Score 102; DB 2; Length 563;

Best Local Similarity 21.1%; Pred. No. 2.2; Mismatches 142; Indels 102; Gaps 16;

Matches 80; Conservative 56; Mismatches 142; Indels 102; Gaps 16;  
 QY 7 TIDPKYIKALKLNQSLSLFQNSQSLQRAVSTPVSYYRIILQKENEKQALAHKICIS 66  
 DB 185 TLPIPIAATIQSLDPVILNFMVNSAGI-----TPLESSRLINLKNLEISDEQYHGLCLP 239  
 QY 67 ILEFFKNLLFVHLISLKNQREGCSTDM-----AVSTPFNNRLMYRLSSRFSLMKS 120  
 DB 240 DKEIIOMLLBRAYATKFRIRGVNTDLCKRSIHVSTLVPLF--QVTEKIGKNTSDLMFA 297  
 QY 121 YCPREFLDVLEAFGLSLDFLDHQAIVKFELETHFSYVPVGFAPHOVYSLQDRYFPI 180  
 DB 298 LC-----EIDGL-----ECVLKY--RPPFIQHDYTGKLP----- 325  
 QY 181 ASVNRITLQKDNFSLTPDLIHDHLPMLHSPSEPFINMGRLFKVIEKQALPSKQ 240  
 DB 326 ---LQCFNDDISYN--FHLLGR---LDDCGVSIYKSHSLTYSKFDKLESIESGLS 376  
 QY 241 RI-----QTIQSLNLIIVRCFW-----FVYESGLINNHGKRAYGAV 277  
 DB 377 LILVLDIEAKFYDPSNEDIQPRYIFLKNVFWTARVNLKYOCFTLDSILDEDE-----T 429  
 QY 278 LISSPQELGHAFIDNVRVLPLELDQIIRLPNTSTPOETLFSIRHFDLVELTSKLEWML 337  
 DB 430 IIGN--LGEESCICQVRL--ISQITIL-----EKRMGLVALLLEIITHALMLAACFR 476  
 QY 338 DQGL-----LESIPLYNOEK 352

DB 477 DKGEVPSDLGDTITLYOER 496

## RESULT 33

C81624  
 exodeoxyribonuclease V, beta chain, probable CP0007 (imported) - Chlamydia pneumoniae  
 C:Species: Chlamydia pneumoniae  
 C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
 C:Accession: C81624  
 R:Read, T.D.; Brubham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hick  
 C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salz  
 Nucleic Acids Res. 28, 1397-1406, 2000  
 A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR3  
 A:Reference number: AB1500; MUID:20150255; PMID:10684935  
 A:Accession: C81624  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1050 <REA>  
 A:Cross-references: GB:AE002164; GB:AE002161; NID:G7188939; PIDN:AA937903.1; PID:G71  
 C:Experimental source: strain AR39, HL cells  
 C:Genetics:  
 A:Gene: CP0007  
 Query Match 5.4%; Score 102; DB 2; Length 1050;  
 Best Local Similarity 22.3%; Pred. No. 5;  
 Matches 78; Conservative 51; Mismatches 116; Indels 104; Gaps 19;  
 QY 46 RIILQKENEKQALAHKICISIL--EFPKNLF--VHLISLKNQREGCSTDMA----- 95  
 DB 137 RLHKKKALHNSQLVHHITNYLKODLWKNVLFQEOFHLLAVRYNTSKHTSLVDKLLA 196  
 QY 96 ----VSTPFNN-----NLWY--RLSSRFSLMKSCYCPFLDYLEA-----FG 134  
 DB 197 SYQPISSYSSRVERLEQISLMHQIYNSLLEIPK---QVFLDQTLAHISGFKKQPS 252  
 QY 135 LLSDFLDHQAIVKFELETH--FSYVPV--GFAVPHQVYSLQDRYFPIASV----- 182  
 DB 253 ILDDL--HHVVDLYNSEHSLSPFKIATFPKRLA-----RYKCAAFVLENNIS 305  
 QY 183 -VNRITLQKDNF-----SLTPDLIHDH--LGHVPMHLHSPSEPFINMGRLFKVIEKQV 235  
 DB 306 WKEITLFCPLDIFPTLVLVDQETLKQNTYF--LSDESVF-----ALEKLLSS 354  
 QY 236 PSKKQRIQTQSLNLIIVRCFWFTVESGLINNHGKRAYGAVLISPOELGH--AFIDN 292  
 DB 355 SEAPVVQAL-----RQYQVLVIDEPDIDKQWISFSN 389  
 QY 293 VRVPLELDQIIRLPNTSTPOETLFSIRHFDLVELTSKLEWMLDQGL 341  
 DB 390 LFISPKFTGSL-----FLIDGPKQSIYEWRSADLPYTLTAKSSPSEDKOL 434

## RESULT 34

S38160  
 NUP133 protein - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: protein YKR082W; protein YKR402  
 C:Species: Saccharomyces cerevisiae  
 C>Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 21-Jul-2000  
 C:Accession: S38160; S42011; S51915; S39123  
 R:Baladron, V.; Ballester, J.P.G.; Bou, G.; del Rey, F.; Estebar, P.F.; Garcia-Cantale  
 submitted to the Protein Sequence Database, March 1994  
 A:Reference number: S38158  
 A:Accession: S38160  
 A:Molecule type: DNA  
 A:Residues: 1-1157 <BAL>  
 A:Cross-references: EMBL:Z28107; NID:G486562; PID:G486563; MIPS:YKR082W  
 A:Experimental source: strain S288C  
 R:Garcia-Cantalejo, J.; Baladron, V.; Estebar, P.F.; Santos, M.A.; Bou, G.; Remacha,  
 Yeast 10, 231-245, 1994  
 A:Title: The complete sequence of an 18,002 bp segment of Saccharomyces cerevisiae ci  
 A:Reference number: S42009; MUID:94262327; PMID:8203164  
 A:Accession: S42011  
 A:Molecule type: DNA

A:Residues: 1-1157 <GAR>  
A:Cross-references: EMBL:D27116; NID:g415899; PIDN:CAA81633.1; PID:g415902  
A:Experimental source: strain S288C  
A:Boye, V.; Wept, R.; Hurt, E.C.  
EMBO J. 13, 6062-6075, 1994  
A:Title: A novel nuclear pore protein Nup133p with distinct roles in poly(A) (+) RNA tran  
A:Reference number: S51915; MUID:95112817; PMID:7813444  
A:Accession: S51915  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-1157 <DOY>  
A:Cross-references: EMBL:X80066; NID:g624033; PIDN:CAA56372.1; PID:g624034  
C:Genetics:  
A:Gene: SGD:NUP133  
A:Cross-references: SGD:S0001790; MIPS:YKR082W  
A:Map position: 11R  
C:Keywords: nucleus; transmembrane protein  
F:413-429/Domain: transmembrane #status predicted <TM1>  
F:641-657/Domain: transmembrane #status predicted <TM2>

Query Match 5.3%; Score 100.5; DB 2; Length 1157;  
Best Local Similarity 21.4%; Pred. No. 7.5;  
Matches 72; Conservative 52; Mismatches 130; Indels 83; Gaps 14;

QY 30 NSOSLQRAYSTPYSVYRILOKENKEKALARHKICISILEFFKNL--LFVHLLSLSKQOR 87  
Db 773 NFINNNYDDNHDWNVLCVKNLKEQ-----CIIAEFYKDLGLVQTLQTDND- 825  
QY 88 EGCSTDMANVSYPFENNLYWLLSSRSLMKSCPRFFLDYLEAGLLSDFL-----DH 142  
Db 826 ---STVSIVET-FPNE-----FPKEFSF-----TLFEYLKHKKLNDLIFRPQOH 868  
QY 143 QAVIKFELETHFSYYPGVGFAVPHQYLSLQDRYFIASVMRTLDKDNFSITPDLIHL 202  
Db 869 DVLIIQPFQSCA-----PKYGHVAMVQ--QILDGSYADAMNVLKNTVDDSKGESL---- 917  
QY 203 LGHPVMLHPSEFPFNNGRUFTKVIKQVALPSKKORIQTLOSNIIAIVRCFWFTVES 262  
Db 918 -----SCEHLNVAKLSSLLVEK-----DNLDDINTLRKIQYNLDTIDA 956  
QY 263 GLIENHEGRKAYGAVLISGPQELGHAFIDNVRVLPLELDQIIRLPFTNTSPQETLFSIRH 322  
Db 957 E--KNTSNLKKGEVQICKRFNGS-----IREVFNLVEELKSTTVN 998  
QY 323 FDELVELTSKLEWMLDQGLLESIPLYNQEKYLSGFEV 359  
Db 999 LSDVELVYSMLD--DBESLFIPLRLLSVDGNLLNFEV 1033

RESULT 35  
T09220  
exocyst complex protein sec5 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: T09220  
R:Keen, Y.; Yoo, J. S.; Hazuka, C.D.; Peterson, K.E.; Hsu, S.C.; Scheller, R.H.  
Proc. Natl. Acad. Sci. U.S.A. 94, 14438-14443, 1997  
A:Title: Subunit structure of the mammalian exocyst complex.  
A:Reference number: Z16617; MUID:98070770; PMID:9405631  
A:Accession: T09220  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-924 <KEB>  
A:Cross-references: EMBL:AF032666; NID:g2827157; PIDN:AA01578.1; PID:g2827158  
A:Experimental source: tissue type brain  
C:Genetics:  
A:Gene: sec5

Query Match 5.2%; Score 99; DB 2; Length 924;  
Best Local Similarity 19.9%; Pred. No. 7.4;  
Matches 90; Conservative 77; Mismatches 122; Indels 164; Gaps 27;

QY 19 KLKQSLSLFQNSQLQRAYSTPYSVYRILOKENKEKQALRHKICISILEFFKNLLFVH 78

```

Db 231 QSLQANOHAFQWON---YAKVLEFNSNYQEAHYHAKKCLSDSHATYQEDLAVLLMR 286
Qy 262 SGLIEN-HEGRKAYGAVLISSPOELGHAFIDNVRVLPLE--LDQIR--LPRNTSTPQ 314
Db 287 MGVDNDFEGKKYKYSRLEFVAHNETFSTYHYNESLAKFKFQVDAFKKREYLVFCEQDFG 346
Qy 315 ETLFSIRHDELVELTSKLEWMLDOGLLESIPLYNOEKYLS 355
Db 347 DTMVYRCLCKICKIASKVLFAFQSAWYEMFK--NQIKFELN 385

```

## RESULT 37

```

C64208
hypothetical protein MG075 - Mycoplasma genitalium
C:Species: Mycoplasma genitalium
C:Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 07-Dec-1999
C:Accession: C64208
R:Praser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;
M.; Finkham, J.; Nguyen, D.; Uterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.
C.A.; Venter, J.C.
Science 270, 397-403, 1995
A:Title: The minimal gene complement of Mycoplasma genitalium.
A:Reference number: A64200; MUID:96026346; PMID:7569993
A:Accession: C64208
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1024 <TIGR>
A:Cross-references: GB:U39697; GB:U43967; NID:G1045744; PID:G1045751; TIGR:MG075
A:Experimental source: strain G-37
C:Genetics:
A:Genetic code: SGC3

```

```

Query Match 5.2%; Score 98.5; DB 2; Length 1024;
Best Local Similarity 19.9%; Pred. No. 9.3;
Matches 68; Conservative 41; Mismatches 134; Indels 99; Gaps 11;

```

```

Qy 29 QNSQLOKAVSTPYSYIILQKENEK--QALARKKCSILEFFKNLLFVHLLSKN 85
Db 564 QOTDLSKLVFSV---IGDILSETNVKIKTLHAVKNELLSLVERASTLTKIKHL----- 613
Qy 86 QREGGSDMAVSTPFRRNLMVRLSSRSLSKSYCPRPFLDYLEAFGLSDPLDQAV 145
Db 614 -----NQYKVLVDKELKNSFIK-----ELNFPDPTKDTIPT 647
Qy 146 IR--FPELETH-----FSYPSGFAVPHOYLSDDRFPPLASVKTLDQNF 192
Db 648 IKKVLFESENKTKLRKKYENEGFPGYMAKFIIVGTFNSAENTFSAL-----DKT 698
Qy 193 SLTPDLIHDLGHVFWMLHPSFSEFFINMGRFLT-----KVLEKVALPSKKRIQT 244
Db 699 KSIRDLFADML-FGKSLESVNDSDSFKIKNGSFTLKAGDNINLTPNHSILTKNVGYOI 757
Qy 245 LOSNLIAIVRCFWFTVESGLIENHEGRKAYGAVLISSPOELG----- 286
Db 758 VVNVNHIDARLTLAELQNTVPSNPK-----PVIKSPVELSKSLFEVWKTIFENSVOI 810
Qy 287 ---HAFIDNVRVLPLELDQIRLPFNTSTPQETLSIRHFD 324
Db 811 LKKEYTFKDNLKFFPFKADGSSRLFDLSKPDQRIVPFAFVD 852

```

## RESULT 38

```

T33079
hypothetical protein F56A6.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 23-Mar-2001
C:Accession: T33079
R:MurRAY, J.; Kohling, T.; O'Neal, D.; Wilson, R.
Submitted to the EMBL Data Library, May 1998
A:Description: The sequence of C. elegans cosmid F56A6.
A:Reference number: Z21279
A:Accession: T33079

```

```

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1846 <MUR>
A:Cross-references: EMBL:AF067217; PIDN:AAC17015.1; GSPDB:GN00019; CESP:F56A6.2
A:Experimental source: strain Bristol N2; clone F56A6
C:Genetics:
A:Gene: CESP:F56A6.2
A:Map position: 1
A:Insertions: 71/1; 165/3; 247/2; 335/2; 356/2; 369/3; 441/3; 497/3; 539/1; 607/3; 636/2
06/1; 1714/3; 1756/2
C:Superfamily: rat myosin heavy chain; myosin motor domain homology; protein kinase

```

```

Query Match 5.2%; Score 98; DB 2; Length 1846;
Best Local Similarity 23.3%; Pred. No. 22;
Matches 71; Conservative 34; Mismatches 92; Indels 108; Gaps 16;

```

```

Qy 9 DPKYIKLKLKLRQSLSEFQNSQLOKAVSTP-----YSYRIILQKENEKQKOLA 60
Db 217 NPKY-----ARLYFQS---KRLGSLPHIFALADVCYHNMMLKIKEN----- 254
Qy 61 RHCISL-----LEFKNLLFVHLLSKNQGEGSDMAVST-----PFRNLMY 108
Db 255 --QCVYISGESGSGKTESINLHMSHLISLSQKSGTSGSTIGQILISAGPVLFAFNAVTLT 312
Qy 109 RLSSRSFLSKSYCPREF-LDYLEAFGLSDPLDQAVIKFPELET-----H 154
Db 313 NNNSRFG-----KFIKINREN-GWVSG-----ANVEIYLLKSRRIIFQTKGERNVH 359
Qy 155 FSYYPVSG-----FVAPHOYLSLDQRYFPPLASVNRLLDKDNFSLTPDLIHDLG 204
Db 360 VFYYLLGADDEBERKKYFLAKPHDYKLYLNQEPALRGV---ERNEF-----DRLR 408
Qy 205 HVPMLHPSFSEFFINMGRFLTQVLEKVALPSKKRIQTLQSNLIAIVRCFWFTVESGL 264
Db 409 HA-----MSSVGCATQQTITGILISAV-----LLGNITYIRKHGYHSDSEGY 452
Qy 265 IENHE 269
Db 453 IENHE 457

```

## RESULT 39

```

S68856
cytochrome P450 2L - spin lobster
N:Contains: oxidoreductase (EC 1.-.-)
C:Species: panulirus argus (spiny lobster)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-May-2000
C:Accession: S68856; S74194
R:James, M.O.; Boyle, S.M.; Tripathi-Rosenthal, H.G.; Smith, W.C.; Greenberg, R.M.; Si
Arch. Biochem. Biophys. 329, 31-38, 1996
A:Title: cDNA and protein sequence of a major form of P450, CYP2L, in the hepatopancreas.
A:Reference number: S68856; MUID:96201120; PMID:8619632
A:Accession: S68856
A:Molecule type: mRNA
A:Residues: 1-492 <JAN>
A:Cross-references: EMBL:U44826; NID:G1304739; PIDN:AAB03106.1; PID:G1304740
A:Experimental source: hepatopancreas
A:Accession: S74194
A:Molecule type: protein
A:Residues: 1-10, 'X', 12-39 <JAN>
C:Genetics:
A:Gene: CYP2L1
C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C:Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; met
F.235-458/domain; cytochrome P450 homology <Cys>
F.436/Binding site: heme iron (Cys) (axial ligand) #status predicted

```

```

Query Match 5.1%; Score 97; DB 1; Length 492;
Best Local Similarity 19.9%; Pred. No. 4.6;
Matches 74; Conservative 56; Mismatches 106; Indels 136; Gaps 18;

```

```

Qy 56 KQALARKKCSILEFFKNLLFVHLLSKNQBSCSDMAVSTPFRRNLMVRLSSRF 115

```



```
Db 79 KTALESKFECDRDPDYTKLF-----GEGNDVGVV-----FSGNV----- 113
Qy 116 SLWKSYPFRFLDYLEAFGL-----LSDFLDHQAVIKFPELETHFSY---YFVS----- 161
Db 114 -MMWOTH-RRFLRLQRLDGLMGKSRLEAAIQAHEAACLVQELKKHTDQMPLPKSLINLAVLN 171
Qy 162 ---GVPAPHQVLSLQD---RYPIASVNRITLDKNFSLTPDLLHDLGHVPWLL--HPS 213
Db 172 VIKLVADHRY--SLQDOEQYFTQLLTNTDMMQGFAL-----NLFLNPLWLLMITPD 223
Qy 214 FSEPIINNGRLFTKVIKQVLSKQKRIQTIQSLNLIIVRCFWFTVESGLIENHGRKA 273
Db 224 FVKNWGVRLRGVCE-----LKDYNKT-----FIKEHQ----- 253
Qy 274 YGAVLISPOELGHAF-----IDNVRVLP----- 298
Db 254 -ATLDPNPKDLDAYLIDQERKEDPLSTWNIETVRVIMDLFGAGTETTSNIRWTL 312
Qy 299 -----ELDQIIRLPFNFTSQETLSIRHPDELVELTSKLEWMLDQGLLESPL----- 347
Db 313 YLMKYPEVQAKIQREIDAAVFRGLTSLSEHKOKLAYFEATIEV--HRIVSLVPLGVSHY 370
Qy 348 YNQEYLSGFVEV 359
Db 371 TNQDTLAGYRL 382

RESULT 40
G86582
exodeoxyribonuclease V, beta [imported] - Chlamydothilla pneumoniae (strain J138)
C:Species: Chlamydothilla pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: G86582
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ie
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; UID:20330349; PMID:10871362
A:Accession: G86582
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1050 <STO>
A:Cross-references: GB:BA000008; NID:98979110; PIDN:BAA98945.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: zwf
C:Superfamily: glucose-6-phosphate dehydrogenase

Query Match 5.1%; Score 97; DB 2; Length 1050;
Best Local Similarity 22.1%; Pred. No. 13;
Matches 77; Conservative 51; Mismatches 117; Indels 104; Gaps 19;

Qy 46 RILOKENKEQALAHKICISIL--BFFKNLLF---VHLLSLKSNQREGCGTDMW----- 95
Db 137 RLIIKNPALTHSQLVHHITNYLKQDLKWNVLFOEQFHLLAVRYNVTSKHTSSLDKLLA 196
Qy 96 ----VSTPFENR-----NLWY-RLLSRSFLWKSYPFRFLDYLEA-----FG 134
Db 197 SYTQPISSYFSSRVERLEQISLWHQIYNSLLEIPK---QVFLDQLTAHISGKKQPS 252
Qy 135 LLSDFLDHQAVIKFPELETH---FSYVPVS-GFVAPHQVLSLQDRYFPITAS----- 182
Db 253 ILDDL--HHFVDLLYTSETHSSLSFSEFKIAETFNFKRLA-----RYKPCAFTVLNMS 305
Qy 183 -VNRITLKDNE-----SLTDPDLHDL-LGHVPWLLHPSSEFFINMGRLTKEVQKQAL 235
Db 306 WVERTLEFCNDRINFNTLLVDLQELKQNTYTW-LSPDESVF-----ALEKLLSS 354
Qy 236 PSKKQRIQTLQSLNLIIVRCFWFTVESGLIENHGRKAYGAVLISSPOELGH---AFIDN 292
Db 355 SBAQVQVQAL-----REQYQLVLIDEFQDTKQOWISFN 389
Qy 293 VRVLPLELDQIIRLPFNFTSQETLSIRHPDELVELTSKLEWMLDQGL 341
Db 390 LFISPKFTGSL-----FLIGDPKQIYEWASDLPTLYLTAKGSFSEDKQL 434
```

```
RESULT 41
F86520
glucose-6-P dehydrogenase [imported] - Chlamydothilla pneumoniae (strain J138)
C:Species: Chlamydothilla pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: F86520
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ie
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; UID:20330349; PMID:10871362
A:Accession: F86520
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-512 <STO>
A:Cross-references: GB:BA000008; NID:98978611; PIDN:BAA98944.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: zwf
C:Superfamily: glucose-6-phosphate dehydrogenase

Query Match 5.1%; Score 96.5; DB 2; Length 512;
Best Local Similarity 18.8%; Pred. No. 5.3;
Matches 83; Conservative 57; Mismatches 124; Indels 177; Gaps 22;

Qy 20 LRQSLSLFFNQSLSQRAY-----STPYSYRIILQENKKEQALAHKICISILE 69
Db 113 LKDSL-----EDLQKTYTGRNRLFVLSLTPPQVFSRIENLNKHK----- 152
Qy 70 FPKNLLFVHLLSLKSNQREGCGTDMAVVSTPF-----FNRNLWY----- 108
Db 153 ----LFV-----KNQDQKPSWRSVRIEKPGRDLDSAKQLQOCINENLNENSVYHIDH 201
Qy 109 -----RLLSRF-----SLWKSYPFRFLDYLE-----AFGLLSDFLDHQAVIK 147
Db 202 YLGKETVQNILITTFANTIFESCWSQ----YIDHVQISLSETIGISRGNFPEKSGMLR 257
Qy 148 -----PFELETHFSYVPVSQVAPH---QYLSLQDRYFP---ASVMR----- 185
Db 258 DMVQNHMMQLLCLLTWEPTTFDAEIRKEKILQ-RISFSEGSIVRGQYGPQTVOG 316
Qy 186 -----TLKDNFSLTPDLIHLGHVPWLLHPSSEFFINMGRLTKEVQKQAL 235
Db 317 VSVLYGREENVDKSRVETVYVAKTVINNPRWLGVP---FYLRAGKRLAKKSTDISII 372
Qy 236 PSKKQRIQTLQSLNLIIVRCFWFTVESGLIENHGRKAYGAVLISSPOELGHAFIDNVR- 294
Db 373 FKKS-----PYNLFAAECSRCPIENDLL-----IIRIQPDE-GVALKFNCKV 414
Qy 295 -----VLPLELDQIIRLPFNFTSQ-----ETLFSIRHFDLV 327
Db 415 PGTNNIVRPVKMDFRYSYFOITTPAYERLLDCIIGDRTLFTGDEVMASWKLFTPV 474
Qy 328 ELTSKLEWMLDQGLLESIPLY 348
Db 475 E-----EW--DQDSFSFPNY 488

RESULT 42
C72103
glucose-6-P dehydrogenase - Chlamydothilla pneumoniae (strain CWL029)
C:Species: Chlamydothilla pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: C72103
R:Kalan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; UID:99206606; PMID:10192388
A:Accession: C72103
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-512 <ARN>
```

A:Cross-references: GB:AE001609; GB:AE001363; NID:g4376502; PID:AA018391.1; PID:g437651  
 A:Experimental source: strain CWL029  
 A:Gene: 2w1  
 C:Superfamily: glucose-6-phosphate dehydrogenase

Query Match 5.1%; Score 96.5; DB 2; Length 512;  
 Best Local Similarity 18.8%; Pred. No. 5.3; Mismatches 124; Indels 177; Gaps 22;  
 Matches 83; Conservative 57;

QY 20 LKOSLSFFONSQSLQRAY-----STPYSYRIILQENKEKQALAHKICISILE 69  
 DB 113 LKDSL-----EDLDKTYGTRGNRLFLYLSTPPQYFSRIIENLNRK----- 152  
 QY 70 FPKNLLFVHLISLKNQREGCSTDMAVSTPF-----FNRNLWY----- 108  
 DB 153 -----LFY-----KNQDQKPMGRVIEKPFGRDIDSAKOLQCCINENLINSYHIDH 201  
 QY 109 -----RLSSRF-----SLWKSYPFRFLDYLE-----AFGLSDPLDHQAVIK 147  
 DB 202 YLGETVQNIILTRFANTIFESCNSQ-----YIDHVISLSETIGIGSKGNPFESGMLR 257  
 QY 148 -----FPELETHESYVVSQFVAPH-----QYLSLDQRYFPI-----ASVWR----- 185  
 DB 258 DMVQNMHMQLCLTMEPTTFDDELREKIKILO-RISPFSEGSISVROQYGRGVQ 316  
 QY 186 -----TLDKNBSLTPTDLIHDLGHVPLHPSSEPFIMKGRFLTKVIEKQAL 235  
 DB 317 VSVLGYREBNVDKDSREYEVVAKTVINPRMLGVF-----FYLRAGKRLAKKSTDSII 372  
 QY 236 PSKKORIQTLOSNIATVRCFWFTVESGLIENHGRKAYGAVLISPOELGHAFIDNVR- 294  
 DB 373 FKKS-----PYNLFAAECSRCPIENDLL-----IIRIQPB-GVALKFNCKV 414  
 QY 295 -----VPLELDQIIRLPENTSTPO-----ETLFSIRHPELV 327  
 DB 415 PGTNNIVRPVKMDPFDYSYFQTTPEAYERILCDCTIGDRTLEFGDEVMASWKLFPVL 474  
 QY 328 ELTSKLEWMLDQGLLESIPLY 348  
 DB 475 E-----EW--DQSSPSFPNY 488

RESULT 43  
 S62468  
 probable membrane transporter - fission yeast (*Schizosaccharomyces pombe*)  
 C:Species: *Schizosaccharomyces pombe*  
 C:Date: 16-May-1996 #sequence\_revision 13-Mar-1997 #text\_change 21-Jan-2000  
 C:Accession: J38579, S62468  
 R:Badcock, K.; Churcher, C.M.; Barrrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
 submitted to the EMBL Data Library, October 1995  
 A:Reference number: Z21745  
 A:Accession: J38579  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-529 <BA2>  
 A:Cross-references: EMBL:Z54354; NID:g1019398; PID:CAA91178.1; PID:g1019411; GSPDB:GN00  
 A:Experimental source: strain 972h-; cosmid c2G11  
 C:Genetics:  
 A:Gene: SPAC2G11.13  
 A:Map position: 1L

Query Match 5.1%; Score 96; DB 2; Length 529;  
 Best Local Similarity 22.0%; Pred. No. 6.1; Mismatches 127; Indels 72; Gaps 11;  
 Matches 66; Conservative 35;

QY 120 SYCPRFPLDYLEAGFLSD--FLDHQAVIKFPELETHESYVVSQFVAPHQLSLDQRY 177  
 DB 133 SWSPSFLGKIVFLVLDNLFISSQCYDSF-LPIFLAFYITR--GPITLESALQDET 189  
 QY 178 FFIASVMRLDKDNFSLSLPDI-HDLIGH-----VPLHPSSEPF-- 218  
 DB 190 DDUDSYITNTTIDSEEPYILHSLILNESAPPAVDDEBHKAKTAARLSVGFOSFPGA 249

QY 219 -IMNGRLFTKVIEKQALPSKKORIQTLOSNIATVRCFWFTVESGL-----IENH 270  
 DB 250 ALFQIIFTRPILKTNND-----ILPFTVVCSCWMLISTPLCTIVLPPVENHSS 302  
 QY 271 RKAYGAVLISPOELGHAFIDNVRVPLELDQIIRLPFN-----TSTPQETLS----- 319  
 DB 303 -DALTLVNSVESHYSFKAMSSISIRLFLRSRLPINCIGTSSSAVIFGKARLNS 361  
 QY 320 -----IRHDELVELTSKLEWMLDQGLLESIPLYNGEYLSGFE 358  
 DB 362 NFQITLLGNGISSFALLGTVIIPLYLTFEYFQLSNQVVMIIISLLPMAPLYGLIGYIPGE 421

RESULT 44  
 H72041  
 exodeoxyribonuclease V, beta - Chlamydia pneumoniae (strain CWL029)  
 C:Species: *Chlamydia pneumoniae*, *Chlamydia pneumoniae*  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000  
 C:Accession: H72041  
 R:Kallan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, Nature Genet. 21, 385-389, 1999  
 A:Title: Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*.  
 A:Reference number: A72000; MUID:99206606; PMID:10192388  
 A:Accession: H72041  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1050 <ARN>  
 A:Cross-references: GB:AE001655; GB:AE001363; NID:g4377039; PID:AA018877.1; PID:g43  
 A:Experimental source: strain CWL029  
 C:Genetics:  
 A:Gene: recB

Query Match 5.1%; Score 96; DB 2; Length 1050;  
 Best Local Similarity 22.1%; Pred. No. 15; Mismatches 117; Indels 104; Gaps 19;  
 Matches 77; Conservative 51;

QY 46 RIILOKENKEKQALAHKICISIL--EFPKNLF--VHLISLKNQREGCSTDMA----- 95  
 DB 137 RLHKNPALHLSQVLVHNTVNLKODLWKVNLFOEPHLLAVRYNTSKHTSLVDKLLA 196  
 QY 96 -----VVSTPEFNR-----NLWY-RLSSRFSLMKSCYCRPFLDYLE-----FG 134  
 DB 197 SYTQPISSYSYSSRVERLEQISLMHQIYNSLLEIPK-----QVPLDQLTHISGFKKQPS 252  
 QY 135 LLSPLDQAVIKFPELETH--PSYVVS-GFVAPRQYLSLDQRYFPIAS----- 182  
 DB 253 ILDDL-HHVDLVYSEHSSLSRPFKIAITRFKRLA-----RYKCAAFVLENNIS 305  
 QY 183 -VWRTLDKDNF-----SLTPDLIDL-LGHVPLHPSSEPFIMKGRFLTKVIEKQAL 235  
 DB 306 WERTLEPCLDRIFNTLVLDQELKONYTPW-LSDBESVF-----ALEKLS 354  
 QY 236 PSKKORIQTLOSNIATVRCFWFTVESGLIENHGRKAYGAVLISPOELGH--AFIDN 292  
 DB 355 SEAPVQAL-----RQYQVLVIDEFOPTDKQWSTFSN 389  
 QY 293 VRVLPLELDQIIRLPENTSTPOETLFSIRHDELVELTSKLEWMLDQGL 341  
 DB 390 LFLSPKFTGSL-----FLIGDPKQSIYEWRSADLPYLTAKSFSFEDKOL 434

RESULT 45  
 T41933  
 large tegument protein - human herpesvirus 7 (strain J1)  
 C:Species: human herpesvirus 7  
 A:Variety: strain J1  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000  
 C:Accession: T41933  
 R:Nicholas, J  
 submitted to the EMBL Data Library, December 1995  
 A:Description: Determination and analysis of the complete nucleotide sequence of hum.  
 A:Reference number: Z22022

A:Accession: T41933  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2059 <N1C>  
A:Cross-references: EMBL:U43400; PIDN:AAC54693.1  
A:Experimental source: strain J1  
C:Genetics:  
A:Note: U31  
C:Superfamily: varicella-zoster virus 240K tegument protein

Query Match 5.1%; Score 96; DB 2; Length 2059;  
Best Local Similarity 20.6%; Pred. No. 37;  
Matches 90; Conservative 62; Mismatches 164; Indels 120; Gaps 23;

QY 13 ILKIALKRLQSLFFQNSOS-LQRAYSTPYSYVRIILQENKEKQALARKK--CISILE 69  
DB 1265 ILKWLIVFVKELNTFFVATMSERGEVVPDYKHFR-ALEYEINSKYIEIKNICNEIIE 1323  
QY 70 FPKNL-LFVHLLSLSKNQREGCSTD----MAVSTPFNNLWYRLLSRPF--SLWKSX 121  
DB 1324 NTDNEKSLTLKIDPNRTAGKQKQFDYLSKILTAETNQO-----QRYKEQKKQY 1377  
QY 122 CRPFELDYLEAGLLSDFLDHQAIV-----KPELETHFSYYP-----VSGF 163  
DB 1378 FD-LLDNIAHFRFAFDNFQNNLILKDKDKFKTLRTDTVERFPNLDLDTFVSSMNVNF 1435  
QY 164 VAPHQYLS-----LLQDRYFPI-ASVMTLTKDNFSL-----TP- 196  
DB 1436 LQALEALSHFVQAQNFQNLVTEQADLPQNTFIPVELSTVKTIKSDINLRMKIHTPO 1495  
QY 197 -----DLIHLGLGHVPMLLHPSSEFFINMGRL-FTKVI--EKVOALPSKKQRI 242  
DB 1496 TPFQVDSVFNTQLIVDEKG-IPQVFNPHNIVFPPFALNVKKIIVPKVLNLVSTKYI 1554  
QY 243 QFQSLNLAIVRCFW-FTVSSGLIENHGR-----KAYGAVLIS-----SPQ 283  
DB 1555 LTTLSILSVKSPKKEINFDLTSYFQKAEFTFQNVFPIINLKIYIITQAWSTSD 1614  
QY 284 ELGHAFIDNVRVPLEDQIIRLPFTSTPQETLSIRHPDELVELSKLEWMLDOGLLE 343  
DB 1615 ETQHSF-----ELPKEFSLIANN-----PEFLF-----GSLQCPDLAINS 1653  
QY 344 SIPLYNQKSLSGREV 359  
DB 1654 LIPLLEKKKYFTPTI 1669

RESULT 46  
E70066  
hypothetical protein ywqB - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
C:Accession: E70066  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berto  
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd  
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle  
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, R.; Masuda, S.; Maues  
Y.M.; Ogawa, K.; Ogata, A.; Oudega, E.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
A.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Segiguchi, J.; Sekowska, A.; Seron  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpsira, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
I.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A:Authors: Yoshikawa, H.F.; Zumschein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: E70066  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-536 <KUN>

A:Cross-references: GB:Z99122; GB:AL009126; NID:G2636029; PIDN:CAB15644.1; PID:G2636152  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: ywqB  
C:Superfamily: Bacillus subtilis hypothetical protein ywqB

Query Match 5.1%; Score 95.5; DB 2; Length 536;  
Best Local Similarity 21.8%; Pred. No. 6.8;  
Matches 57; Conservative 30; Mismatches 93; Indels 81; Gaps 12;

QY 107 WYRLLSRFRSLWKSXCYPR-----PFLLDYLEAF-----GLLSDFLDHOAV---IKF 148  
DB 142 WLAFDSEFSLWQARTPEGSONQGLYGYLSALKKHAPKPELKSLYQIHSIAVCLRM 201  
QY 149 FELETHFSYYPVSGFVAPHQYLSLLQDRYFPFIASVMT-----LQKDNFSLTPDL 198  
DB 202 FTLTEAGKLNEQDFSLNPYVQQLMDTIYSSIDKLTVALYSPALDFLDK-----TPDV 256  
QY 199 IHDLLGHVPMLLHPSSEFFINMGRLFTKVIKVOALPSKKQRIQTLOSNLIAIVRCFWF 258  
DB 257 IRLHL-----LKEIFQYERI-----RVFGEIWSALLSRPK-----WV 289  
QY 259 TVESGLIENHGRKAYGAVLISPOELGHAFID---NVRVLPLELDQIIRLPFTSTPQ 314  
DB 290 AREQIILKKEAGRR-----FSPQLQFGLHLEFLQKNDVIFBEADQF-----PP 334  
QY 315 ETLFSIRHFDVELVTSKLEW 335  
DB 335 EAL--PYTFQWLSEMTAKDW 353

RESULT 47  
B84846  
probable zinc proteinase [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: B84846  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bonito, M.I.; Town, C.D.; Putil, C.V.; A  
eues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: B84420; MUID:20083487; PMID:10617197  
A:Accession: B84846  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-970 <STO>  
A:Cross-references: GB:AE002093; NID:G2335108; PIDN:AAC02769.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g41790  
A:Map position: 2  
C:Superfamily: inaeulysin

Query Match 5.0%; Score 95; DB 2; Length 970;  
Best Local Similarity 21.7%; Pred. No. 17;  
Matches 97; Conservative 48; Mismatches 128; Indels 174; Gaps 23;

QY 9 DPXYILKALKRQSLSPQNSOSLQRA-----YSTPYSYVRIILQENKEKQALAR 61  
DB 421 NPAIVQKVDELSPSNFRIFWESQKQEGQTDKAEFWYNTAISLEKITSS-----469  
QY 62 HKCISILSFFKNLLFVHL-----LSLSKQREGCSTDMAVSTPFNNLWYR 109  
DB 470 ---TIQWVQASPDVHLHLPAPNVIPDLSL-KDADDKETVPVLURKTPF---SRLWYK 522  
QY 110 LLSRFRSLWKSX-----CP-----RFFLDYLEAF-----133  
DB 523 -PDMFSKPKAYKMDPNCPLAVSSPDAAVLDTIFRLMDYLNAYAYAAQVAGLYGVS 581  
QY 134 -----GLLSDF---LDHOAVIKFFFELETHFSYYPVSGFVA 165  
DB 582 LSDNGFELTLGYNHKLRLILLETVVGVKIANFENKPDRAVIK---ETVTKEYQNYKFRQ 637

```

QY 166 P-HO---YLST-LDRFPIASVWRTLDKDNFSLTPDLIHLGLH-----VPMILH 211
DB 638 PTHQMMYCSLILDDQTP-----WTEL--DVLSHLEDAVKYFVLLS 681
QY 212 PFSSEFFI-----NMGRLFTKVIKY-----QALPSKQRIQTLOSNTLAIYRC 255
DB 682 RTFECYIAGVNNENAEASVYKHLIEDVLENDPKRPLPFS-----QHLTRVVKL--- 733
QY 256 FWFVTEGSLIENHEGRKAYGAVLISFQELGHAFLDNVAVLPLELDQIIRLP-FNTSTPQ 314
DB 734 -----GEGMKYFYHQDSNPSDENSAIVHYIQVHRDPSNMKIKQLFGLVAKQ 781
QY 315 ETLPSIRHDELVELVETSKLEWMLDOGL 341
DB 782 ATFHQLRTVEOLGYITA-LAQRNDSGI 807

RESULT 48
S8B8 protein - Yeast (Saccharomyces cerevisiae)
N.Alternate names: cyclin-dependent kinase 5 activator homolog; probable membrane protei
C.Species: Saccharomyces cerevisiae
C.Date: 19-Mar-1997 #sequence revision 25-Apr-1997 #text change 20-Jun-2000
C.Accession: S74293, S19496, S19495, S19766; C57062; S53594; S55868; S68611; S59988
R.Wedder, H.; Wandut, R.
Submitted to the Protein Sequence Database, September 1996
A.Reference number: S74288
A.Accession: S74293
A.Molecule type: DNA
A.Residues: 1-1427 <MED>
A.Cross-references: EMBL:X59720; NID:G1907116; PID:G1907218; MIPS:YCR081W
R.Ballesta, J. G.; Franco, L.; Hoenicka, J.; Jimenez, A.; Remacha, M.; Sanz, E.
Submitted to the Protein Sequence Database, March 1992
A.Reference number: S19486
A.Accession: S19496
A.Molecule type: DNA
A.Residues: 202-863 <BAL>
A.Cross-references: EMBL:X59720; MIPS:YCR080W
A.Accession: S19495
A.Molecule type: DNA
A.Residues: 531-1427 <FEL>
A.Cross-references: EMBL:X59720; MIPS:YCR080W
R.Hengartner, C.U.; Thompson, C.M.; Zhang, J.; Chao, D.M.; Liao, S.M.; Koleske, A.J.; Ok
Genes Dev. 9, 897-910, 1995
A.Title: Association of an activator with an RNA polymerase II holoenzyme.
A.Reference number: A57062; MUID:95293223; PMID:7774808
A.Accession: C57062
A.Status: not compared with conceptual translation
A.Molecule type: DNA
A.Residues: 202-1427 <HEN>
R.Jimenez, A.
Submitted to the EMBL Data Library, December 1992
A.Reference number: S53589
A.Accession: S53594
A.Molecule type: DNA
A.Residues: 1-133, 'RFV', 136-137, 'FDMS', 142, 'LASLYLTA' <JIM>
R.Rodriguez-Cousino, N.; Lill, R.; Neupert, W.; Court, D.A.
Yeast 11, 581-585, 1995
A.Title: Identification and initial characterization of the cytosolic protein Ycr7p.
A.Reference number: S55867; MUID:95373282; PMID:7645349
A.Accession: S55868
A.Molecule type: DNA
A.Residues: 1-133, 'RFV', 136-137, 'FDMS', 142, 'LASLYLTA' <ROD>

```

```

A.Cross-references: EMBL:X59720
R.Huang, Q.Q.; Lee, K.Y.; Wang, J.H.
FEMS Lett. 378, 48-50, 1996
A.Title: A novel yeast protein showing specific association with the cyclin-dependen
A.Reference number: S68611; MUID:96140706; PMID:8549800
A.Accession: S68611
A.Molecule type: protein
A.Residues: 701-1011, 1013-1056 <HUA>
A.Genetic:
A.Gene: SGD:S8B8
A.Cross-references: SGD:S0000677; MIPS:YCR080W; MIPS:YCR081W
A.Map position: 3R
C.Superfamily: Saccharomyces cerevisiae S8B8 protein
C.Keywords: nucleus; transmembrane protein
F:277-293/Domain: transmembrane #status predicted <TM1>
F:353-369/Domain: transmembrane #status predicted <TM2>
F:566-582/Domain: transmembrane #status predicted <TM3>
F:693-709/Domain: transmembrane #status predicted <TM4>
F:724-740/Domain: transmembrane #status predicted <TM5>
F:1002-1018/Domain: transmembrane #status predicted <TM6>
F:1081-1097/Domain: transmembrane #status predicted <TM7>
F:1169-1185/Domain: transmembrane #status predicted <TM8>

Query Match 5.0%; Score 95; DB 2; Length 1427;
Best Local Similarity 18.8%; Pred. No. 28;
Matches 98; Conservative 65; Mismatches 155; Indels 204; Gaps 25;

QY 12 YLTKIALKLRQSLSLFPQNSGLRAYSPTSYRRI-----LQENKEQALAHKCI 65
DB 214 YLLE---KLFDMTHNYNDSQOL-RTWKROLSYFKLGNCSYLRINNE---IFHMLV 266
QY 66 SLIEFFKNLLF-----VHLISLKNQREGCSTDMAVST-----PFF----- 102
DB 267 EFINKMENFEFLPLSLHILMIFWMDICQIDINAVAAITITSQKEPFLVTKITDMLHK 326
QY 103 -----NRNLVYRLSSRSFSL-----WKSS 120
DB 327 YVIVSSKSMINDENYIINDIKNNKIKLNTKLISLILKIFQEOSLEVFIFPTSNMEI 386
QY 121 YCRPFF-----LDYLEAFGLSDFLDHOAVIKFPELETHESYVPGFVAP 166
DB 387 KYPFLFEIYSNADPTQNSDMKKGLLISYRNBSLKNSSIR-----NIVMASNAN 437
QY 167 HQYSLIDRYPPIASVW-----RTLDKDN-----FSLTPDLIHLGLHGP 207
DB 438 DFLITVTCQKPKLSCLQNCIDTQFTKLD-DNPEFEPMPYVQNPFLTHKIIQLIL 496
QY 208 WLLHPS--PSEFIMNGRLFTKVIKYQALPSKQRIQTLOSNTLAIYRCFWTVESG 263
DB 497 WSHPSKQPDHYSN-----QVAKLLL-----KINSDEDLHEFOEDALWSLVFO 544
QY 264 LIENHEGRK-----AYGAV-----LISSP-----QELGHAFLD-- 291
DB 545 LAKNPSAQKRVSYMPSLYRLNLITVYGIKYPIYIRKLISGLIYQDSNDKRVHQ 604
QY 292 ---NVRVLPLELDQIIRLPFNTSTPQETLSIRHDELVELTSKL----- 333
DB 605 LLINLKISPLMSQYNNVLRNMEYDKYEIFPFDDLVITQIKRILSNDITNLQLS 664
QY 334 -----EWMLD--QGLESIRPLYNQEKYLSGFEVLQ 361
DB 665 KTLPSIKIMVAEWYLSHLCSGILSSV--NRTVLKIKFIKFC 703

RESULT 49
A60026
cell communication-mediating membrane protein pTra2a - Caenorhabditis elegans
C.Species: Caenorhabditis elegans
C.Date: 03-Dec-1993 #sequence revision 18-Nov-1994 #text change 08-Oct-1999
A.Accession: A60026
R.Kimabara, P.E.; Okuma, P.G.; Kimble, J.
Mol. Biol. Cell 3, 461-473, 1992
A.Title: tra-2 encodes a membrane protein and may mediate cell communication in the

```

A:Reference number: A60026; MUID:92360913; PMID:1498366

A:Accession: A60026

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1475 <KUM>

A:Cross-references: GB:S42187; NID:9253436; PIDN:AAB2845.1; PID:9253437

A:Note: sequence extracted from NCBI backbone (NCBIN:110873; NCBI:P110874)

C:Keywords: transmembrane protein

Query Match 5.0%; Score 95; DB 2; Length 1475;

Best Local Similarity 16.7%; Pred. No. 29;

Matches 75; Conservative 65; Mismatches 120; Indels 188; Gaps 19;

QY 1 VHYCERTLDPKYILKALKLROS-----LSLFFONSQSLQAYSTPYSYRII 48

DB 77 LHTQHYLLNNISNLKIKFRQTYTLNDVCFKPHITIFQSSSDQNEY--PHYIQRLL 134

QY 49 LQENKEQALARKHCISILEFPKNLLFVH--LLSLSKNQREGCSTDMVAVVSTPFFNRL 106

DB 135 L-----ENQRLSPCLIVTPLNCFYDIYRIHGISNWNKN-----TDFLNR-- 174

QY 107 MYRLSSRFLSKSYCPRPFLDYLEAFG-----LLSDFLDHQAVIKFFEL 151

DB 175 --RLRNS-----YIEAIGENDERPYVKSNGPSLIKSWADH-----MFDL 212

QY 152 ETHFSYYPVSGF-----VAPHOY-----LSLQDRYPPIASV 183

DB 213 -----PSKSFNTSKDALFQKIKLWLLSIEPRQKTCASIIHSCDTPLDSEHYFNICTD 265

QY 184 METLDK-----DNPSLTPDLIHLGHVPLLLHPSSEFFINMGRLETKVI 229

DB 266 MGSVDNFAEKKTEKLEDDVEEPAMMLDCVDQEQFIEWMOE-----LEIKMYSHVT 318

QY 230 EK-----VQALPSKKQRIQTLOSNIIV----- 253

DB 319 EKPDPYVNVNQCDKIFHDUNSTGIEFFPGSFSSTKSDQDTMOTIEVILLTPMELLSAM 378

QY 254 -----RCFWFTVESGLIENHEGRKAYGAVLISSPQELGHAFIDN-----VRVPLE 299

DB 379 QHSDPVGFSIW-TIEKABELIHEPLAL-----KEETERKENRMSKMRVTSRV 429

QY 300 LDQIIRLPNTSTPQETLFSIRHDELV 327

DB 430 LDNTVTTKLOSFSEKQTIHFVNVVHSLI 457

RESULT 50

C81321

Probable cation-transporting ATPase Cj1161c [imported] - Campylobacter jejuni (strain NC

C:Species: Campylobacter jejuni

C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002

C:Accession: C81321

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barral

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp

A:Reference number: A81250; MUID:20150912; PMID:10688204

A:Accession: C81321

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-699 <PAR>

A:Cross-references: GB:AL139077; GB:AL111168; NID:96968444; PIDN:CAB73415.1; PID:G696859

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: Cj1161c

C:Superfamily: Thermotoga maritima P-type cation-transporting ATPase; ATPase nucleotide-

F:8-37/Domain: heavy-metal-associated homology <HMA>

F:519-657/Domain: ATPase nucleotide-binding domain homology <ATN>

Query Match

5.0%; Score 94.5; DB 2; Length 699;

Best Local Similarity 22.5%; Pred. No. 12;

Matches 76; Conservative 54; Mismatches 101; Indels 107; Gaps 18;

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2003, 12:09:28 ; Search time 25 Seconds  
(without alignments)  
600.577 Million cell updates/sec

Title: US-09-438-185A-1047

Perfect score: 1889

Sequence: 1 VHYCETLDPKYLKIALKL.....ESTPLNQEKYLSGFEVLQC 362

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1886	99.8	362	1	AAAH_CHLPN
2	236.5	13.6	289	1	PHAH_VIBCH
3	235.5	12.5	262	1	PHAH_PSEAE
4	231	12.2	445	1	TRSH_CHICK
5	230.5	12.2	444	1	TRSH_HUMAN
6	229.5	12.1	491	1	TY3H_PHASP
7	227.5	12.0	444	1	TRSH_RAT
8	226.5	12.0	447	1	TRSH_MOUSE
9	225.5	11.9	444	1	TRSH_RABIT
10	225.5	11.9	444	1	TRSH_ANGAN
11	225.5	11.9	498	1	TY3H_MOUSE
12	223.5	11.8	481	1	TRSH_XENLA
13	223	11.8	498	1	TY3H_RAT
14	221.5	11.7	490	1	TY3H_BOVIN
15	221	11.7	275	1	PHAH_RHILO
16	220	11.6	528	1	TY3H_HUMAN
17	219	11.5	508	1	TY3H_DROME
18	216.5	11.5	456	1	TY3H_SCHMA
19	214	11.3	453	1	PHAH_RAT
20	213.5	11.3	452	1	PHAH_HUMAN
21	212.5	11.2	453	1	PHAH_MOUSE
22	201	10.6	313	1	PHAH_RALSO
23	196.5	10.4	297	1	PHAH_CHRVO
24	194	10.3	452	1	PHAH_DROME
25	189.5	10.0	294	1	PHAH_CAUCR
26	178.5	9.4	524	1	TY3H_CAEEL
27	175.5	9.3	457	1	PHAH_CAEEL
28	106.5	5.6	1501	1	NINC_DROME
29	104.5	5.5	2331	1	RRPL_MABVP
30	102	5.4	563	1	YAS8_SCHPO
31	100.5	5.3	1157	1	N133_YEAST
32	100.5	5.3	2212	1	RRPL_EBOMZ
33	98.5	5.2	1024	1	Y075_MYCGE

34	97	5.1	492	1	CPL1_PANAR
35	97	5.1	1050	1	EX5B_CHLPN
36	96.5	5.1	512	1	G6PD_CHLPN
37	96	5.1	529	1	YABD_SCHPO
38	95.1	5.1	2059	1	TEGU_HSV7J
39	95.5	5.1	1398	1	TOP2_PLAFK
40	95	5.0	1427	1	SRB8_YEAST
41	95	5.0	1475	1	TRA2_CAEEL
42	93.5	4.9	650	1	Y096_MYCGE
43	92.5	4.9	466	1	Y065_MYCGE
44	92	4.9	447	1	FLP_KIULA
45	92	4.9	526	1	MATK_ARATH
46	92	4.9	544	1	MATK_MAIZE
47	92	4.9	4563	1	APB_HUMAN
48	91.5	4.8	293	1	SPEE_METJA
49	91	4.8	2183	1	RRPL_MEASA
50	91	4.8	2183	1	RRPL_MEASE
51	90.5	4.8	616	1	ORC3_YEAST
52	90.5	4.8	1102	1	TR13_YEAST
53	90.5	4.8	1102	1	TR13_YEAST
54	90.5	4.8	2331	1	RRPL_MABVM
55	90	4.8	835	1	VP3_ROTSL
56	90	4.8	863	1	APCE_GALSU
57	90	4.8	892	1	YLB3_CAEEL
58	90	4.8	4590	1	FATH_HUMAN
59	89.5	4.7	2136	1	YCF2_MARPO
60	89	4.7	1037	1	YQJ8_YEAST
61	89	4.7	1197	1	EVGS_ECOLI
62	89	4.7	1197	1	EVGS_ECOLI
63	88.5	4.7	519	1	ARR5_HUMAN
64	88.5	4.7	891	1	DPH1_HELPY
65	88.5	4.7	1967	1	YG50_YEAST
66	88	4.7	439	1	Y204_METJA
67	88	4.7	659	1	CU4A_HUMAN
68	88	4.7	873	1	YAJ4_YEAST
69	88	4.7	2291	1	R3PB_BEV
70	87.5	4.6	1319	1	CLT1_RAT
71	87.5	4.6	454	1	W2YE_YERPE
72	87.5	4.6	542	1	MATK_ORISA
73	87.5	4.6	594	1	PEK_CANJE
74	87	4.6	324	1	CBH_LACEL
75	86.5	4.6	495	1	VNS3_ROTSP
76	86.5	4.6	986	1	EP1B_STARP
77	86.5	4.6	1202	1	RPW2_YEAST
78	86.5	4.6	1492	1	AT7A_RAT
79	86	4.6	523	1	RPN3_YEAST
80	86	4.6	1405	1	DPOA_SCHPO
81	86	4.6	1845	1	YM68_CAEEL
82	85.5	4.5	839	1	TLR4_HUMAN
83	85.5	4.5	839	1	TLR4_PANPA
84	85.5	4.5	2039	1	CCH1_YEAST
85	85	4.5	238	1	Y106_METJA
86	85	4.5	462	1	SYTM_YEAST
87	85	4.5	493	1	Y100_STRPN
88	85	4.5	1235	1	RNT1_ARATH
89	84.5	4.5	585	1	YHO9_YEAST
90	84.5	4.5	719	1	YB62_CAEEL
91	84.5	4.5	1070	1	EX5C_BUCAL
92	84.5	4.5	1748	1	POLR_ELV
93	84.5	4.5	3668	1	POLG_PEMVC
94	84	4.4	240	1	YC67_AQUAE
95	84	4.4	501	1	PIPD_LACLA
96	83.5	4.4	423	1	ARGD_YEAST
97	83.5	4.4	450	1	W2YE_ECO57
98	83.5	4.4	450	1	W2YE_ECOLI
99	83.5	4.4	614	1	S6AC_RABIT
100	83.5	4.4	835	1	UBPQ_MOUSE
101	83	4.4	396	1	RT09_HUMAN
102	83	4.4	610	1	ABC1_SCHPO
103	83	4.4	652	1	MX1_RAT
104	83	4.4	764	1	DPO3_SULSO
105	83	4.4	962	1	YBX7_SCHPO
106	83	4.4	969	1	YEAC_SCHPO

Q27712	panulirus a
Q92797	chlamydia p
Q92806	chlamydia p
Q09812	schizosacch
P52362	human herpe
P41001	plasmodium
P25648	saccharomyc
P34709	caenorhabdi
P47342	mycoplasma
P47311	mycoplasma
P13783	kluyveromyc
P56784	arabidopsis
P48190	zea mays (m
P04114	homo sapien
Q57761	methanococc
P13576	measles vir
P57247	buchnera ap
P54790	saccharomyc
Q03660	saccharomyc
P31352	marburg vir
P15736	simian 11 r
P35911	galdieria s
P46578	caenorhabdi
Q14517	homo sapien
P09975	marchantia
Q12496	saccharomyc
P58402	escherichia
P30855	escherichia
Q12774	homo sapien
P56105	helicobacte
P53327	saccharomyc
Q06260	methanococc
Q13619	homo sapien
P40309	saccharomyc
P18458	berne virus
Q92428	rattus norv
Q82451	yersinia pe
P12175	oryza sativ
Q9pmu0	campylobact
Q06115	lactobacilli
P35425	simian 11 r
P30195	staphylococ
Q02773	saccharomyc
P70705	rattus norv
P40016	saccharomyc
P28040	schizosacch
P34529	caenorhabdi
O00206	homo sapien
Q9ttu0	pan paniscu
P50077	saccharomyc
Q57570	methanococc
P07236	saccharomyc
Q97p44	streptococc
Q9fjro	arabidopsis
P38818	saccharomyc
Q09437	caenorhabdi
P57528	buchnera ap
P35928	erysinum la
Q01500	p genome po
Q67105	quifex aeo
P49022	lactococcus
P18544	saccharomyc
Q8xaq5	escherichia
P27835	escherichia
P48055	oryctolagus
Q99mx1	mus musculus
P82933	homo sapien
Q92338	schizosacch
P18588	rattus norv
P95979	sulfolobus
Q10201	schizosacch
Q14077	schizosacch

```

107      83      4.4      1014      1      CINI_YEAST      P40987 saccharomyc
108      83      4.4      1254      1      DROC_YEAST      P15801 saccharomyc
109      82.5      4.4      4455      1      ORCS_SCHPO      O41114 saccharomyc
110      82.5      4.4      5442      1      MATK_WHEAT      P58271 triticum ae
111      82.5      4.4      545      1      YB90_YEAST      P58346 saccharomyc
112      82.5      4.4      1001      1      MPD_HELPI      O91357 heliobacte
113      82.5      4.4      1491      1      RT7A_MOUSE      Q64530 mus musculu
114      82.5      4.4      2210      1      RREP_EOSM      Q68802 ebola virus
115      82      4.3      365      1      M2L2_LACLC      P50179 lactococcus
116      82      4.3      365      1      RABG_BACSU      P08836 gallus gall
117      82      4.3      367      1      FPFS_CHICK      Q96r24 rattus norv
118      82      4.3      475      1      ATXX_RAT      P17158 hordium vul
119      82      4.3      641      1      MATK_HORVU      P50831 bacillus su
120      82      4.3      775      1      YPUA_BACSU      P52363 human herpe
121      82      4.3      819      1      YCXA_CHLVU      P56370 chlorella v
122      82      4.3      1104      1      CARB_ANASP      Q8YQ12 anabaena sp
123      82      4.3      1458      1      CUTA_SCHPO      Q9URV2 schizosacch
124      82      4.3      1738      1      YCF1_EPIVI      Q00383 epifagus vl
125      82      4.3      618      1      ORC2_DROME      O24168 drosophila
126      81.5      4.3      915      1      KCQ4_YEAST      P25389 saccharomyc
127      81.5      4.3      2294      1      YCF2_ARATH      P56786 arabidopsis
128      81.5      4.3      3056      1      ATM_HUMAN      Q13315 homo sapien
129      81.5      4.3      276      1      SPEE_BACSU      P70998 bacillus su
130      81      4.3      298      1      ERA_XYUPA      Q9Pb97 xyloella fas
131      81      4.3      323      1      UME3_YEAST      P47821 saccharomyc
132      81      4.3      438      1      RUMA_HAEN      P46643 haemophilus
133      81      4.3      475      1      ATXX_MOUSE      P26638 mus musculu
134      81      4.3      506      1      SYE_CHLIR      O84451 chlamydia t
135      81      4.3      794      1      YE14_YEAST      P33961 saccharomyc
136      81      4.3      880      1      YMD5_CABEL      P34459 caenorhabd1
137      81      4.3      1016      1      FDOG_ECOLI      P32176 escherichia
138      81      4.3      3830      1      SACS_MOUSE      Q9J1C8 mus musculu
139      80.5      4.3      346      1      TRAJ_LESBO      Q48514 leptospira
140      80.5      4.3      377      1      PFTA_RAT      Q04631 rattus norv
141      80.5      4.3      787      1      ECCE_HUMAN      O60344 homo sapien
142      80.5      4.3      826      1      TURA_PAPAN      Q9SP24 homo anub1
143      80.5      4.3      959      1      VPI2_BTIVV      Q06998 bluetongue
144      80.5      4.3      1234      1      YNX5_CABEL      P34578 caenorhabd1
145      80.5      4.3      1365      1      KRES_YEAST      P22023 saccharomyc
146      80.5      4.3      2493      1      YBA4_YEAST      P35194 saccharomyc
147      80      4.2      253      1      VG23_HSVSA      Q01006 herpesvirus
148      80      4.2      272      1      PK1_NPVAC      P41415 autographa
149      80      4.2      325      1      YAB6_YEAST      P31379 saccharomyc
150      80      4.2      325      1      YAB6_YEAST      P31379 saccharomyc

```

## ALIGNMENTS

```

RESULT 1
AAAH_CHLPN STANDARD; PRT; 362 AA.
AC Q9ZEL3;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Probable aromatic amino acid hydroxylase (EC 1.14.16.-).
GN CPN1046 OR CP0806 OR CPJ1046.
OS Chlamydia pneumoniae (Chlamydiophila pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydiophila.
OX NCBI_TaxId=83558;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CM1029;
RA MEDLINE=9920606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Gilmord J., Davis R.W., Stephens R.S.,
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RA MEDLINE=20150255; PubMed=10684935;

```

```

RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uetreback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=2033049; PubMed=10871362;
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CM1029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -1 - COFACTOR: Binds 1 ferrous ion (By similarity).
CC -1 - SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001685; AD19183.1; -
DR EMBL; AE002240; AA73705.1; -
DR EMBL; AP002548; BAA9253.1; -
DR TIGR; CP0806; -
DR InterPro; IPR001273; Aaa_hydroxylase.
DR Pfam; PF00351; Bioplerin_H; 1.
DR PRINTS; PR00372; FYWHYDRXLASE.
DR TIGRFAMs; TIGR01267; Phe4hydrox mono; 1.
DR PROSITE; PS00367; BIOPTERIN HYDROXYL; 1.
KW Oxidoreductase; Monooxygenase; Iron; Complete proteome.
FT METAL 200
FT METAL 205
FT METAL 205
FT SEQUENCE 362 AA; 42513 MW; 01B89B84B4FE593B CRC64;
SQ
Query March 99.8%; Score 1886; DB 1; Length 362;
Best Local Similarity 99.7%; Pred. No. 2e-141; 0; Indels 0; Gaps 0;
Matches 361; Conservative 1; Mismatches
QY 1 VHYCERTLDPKYILKIALKLRQSLSPFONOSQIOPAYSPRYVYRIILQENKEROA 60
DB 1 MHICERTLDPKYILKIALKLRQSLSPFONOSQIOPAYSPRYVYRIILQENKEROA 60
QY 61 RHKISTILFFPKNLVPHLLSLSKNQRGCGTDAAVSTPPNNMLYRLLSRFLWMS 120
DB 61 RHKISTILFFPKNLVPHLLSLSKNQRGCGTDAAVSTPPNNMLYRLLSRFLWMS 120
QY 121 YCPFFFLDYLEAFGLSDFLDHQAIVKPFLETHFSYVPGFAVPHQVYLSLLQDRYPI 180
DB 121 YCPFFFLDYLEAFGLSDFLDHQAIVKPFLETHFSYVPGFAVPHQVYLSLLQDRYPI 180
QY 181 ASVVRTLDKQNFSLTPDLIHDLGLHVPMLLHPSSEFFIMNGRLFTVIEKVALPSKKQ 240
DB 181 ASVVRTLDKQNFSLTPDLIHDLGLHVPMLLHPSSEFFIMNGRLFTVIEKVALPSKKQ 240
QY 241 RIOTLQSNLAIRYRCFWFVESGLLENHGRKAGAVYISPOELGAFFIDNVVPLPL 300
DB 241 RIOTLQSNLAIRYRCFWFVESGLLENHGRKAGAVYISPOELGAFFIDNVVPLPL 300
QY 301 DQIRLPEFNSTQETLSIRHPDELVELTSKLEMLDOGLLSRPLVNOEKLSTFVTL 360
DB 301 DQIRLPEFNSTQETLSIRHPDELVELTSKLEMLDOGLLSRPLVNOEKLSTFVTL 360
QY 361 CQ 362
DB 361 CQ 362

```







CC -!- COFACTOR: FERROUS ION.  
 CC -!- PATHWAY: THIS IS THE RATE-LIMITING ENZYME IN THE BIOSYNTHESIS OF  
 CC SEROTONIN IN THE CENTRAL NERVOUS SYSTEM AND CATALYZES THE FIRST  
 CC STEP OF THE SYNTHESIS OF MELANOTIN IN THE PINEAL GLAND.  
 CC -!- SUBUNIT: MULTIMER OF IDENTICAL SUBUNITS.  
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOPFORMS; 1 (SHOWN HERE) AND 2; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -!- TISSUE SPECIFICITY: ISOPFORM 2 SEEMS TO BE LESS WIDELY EXPRESSED  
 CC THAN ISOPFORM 1.  
 CC -!- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID  
 CC HYDROXYLASES FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X52836; CAA37018.1; -;  
 CC EMBL; L29306; AAA67050.1; -;  
 CC EMBL; AF057280; AAC69459.1; -;  
 CC PIR; S10489; S10489.  
 CC HSSP; P04177; ITOH.  
 CC Genew; HGNC:12008; TPH.  
 CC MIM; 191060; -;  
 CC InterPro; IPR002912; ACT.  
 CC InterPro; IPR001273; Aaa\_hydroxylase.  
 CC Pfam; PF00351; bipterin\_H; 1.  
 CC Pfam; PF01842; AC1; 1.  
 CC PRINTS; PR00372; FWHYDRXLASE.  
 CC PROSITE; TIGR01270; Trp\_5\_monoox; 1.  
 CC Oxidoreductase; Monooxygenase; Serotonin biosynthesis; Iron;  
 CC Phosphorylation; Alternative splicing.  
 CC MOD RES 58 58 PHOSPHORYLATION (BY PKA) (POTENTIAL).  
 CC METAL 272 272 IRON (BY SIMILARITY).  
 CC METAL 277 277 IRON (BY SIMILARITY).  
 CC METAL 317 317 IRON (BY SIMILARITY).  
 CC VARSPLIC 438 444 VARKESI -> SLNEDVLQSVFALLFLFLSLHGCHPDT  
 CC  
 CC CONFLICT 19 19 T -> S (IN REF. 1).  
 CC CONFLICT 68 68 T -> I (IN REF. 1).  
 CC CONFLICT 90 91 TP -> NL (IN REF. 1).  
 CC CONFLICT 97 97 M -> L (IN REF. 1).  
 CC CONFLICT 100 100 E -> D (IN REF. 1).  
 CC CONFLICT 104 104 S -> T (IN REF. 1).  
 CC CONFLICT 151 151 L -> S (IN REF. 2).  
 CC CONFLICT 154 154 N -> S (IN REF. 2).  
 CC CONFLICT 157 157 H -> Y (IN REF. 2).  
 CC CONFLICT 179 179 R -> Q (IN REF. 2).  
 CC CONFLICT 207 207 R -> I (IN REF. 2).  
 CC CONFLICT 217 217 V -> I (IN REF. 2).  
 CC CONFLICT 344 344 A -> V (IN REF. 2).  
 CC CONFLICT 414 414 S -> N (IN REF. 2).  
 CC CONFLICT 419 419 S -> N (IN REF. 2).  
 CC CONFLICT 425 425 Q -> R (IN REF. 2).  
 CC CONFLICT 436 436 A -> G (IN REF. 2).  
 CC SEQUENCE 444 AA; 51004 MW; 86C398869AB8120A CRC64;  
 CC  
 CC Query Match 12.2%; Score 230.5; DB 1; Length 444;  
 CC Best Local Similarity 27.8%; Pred. No. 6.4e-11;  
 CC Matches 64; Conservative 43; Mismatches 96; Indels 27; Gaps 4;  
 CC  
 CC QY 107 WYRLSSRFLSKYCRPFLLDYLEAFGLSLDFLDH-----QAVIKFPELETHFSY 158  
 CC Db 174 WGVTFRELKLYPTHACK---EYLNKPLLSKYCYGREDNIPQLEDVSNFLKERTGFSIR 230  
 CC QY 159 PVSGFVAPQVLSLQRYPIASVMTLDKNFSLTDLHLGLGHVPLLLHPSSEFF 218  
 CC Db 231 PVAGYLSRDLGLAFVHCTQYVHSSDFYFPEPTCHELLGHVPLLABSPAQFS 290

QY 219 INMGRFTKVIKVOALPSKKORIOTLOSNIATVRCFWFTVESGLIENHEGRKAYGAVL 278  
 Db 291 QEIG-----LASLGASEEAVQKLAT-----CVYFTVEGLCKDQGLRVFGAGL 334  
 QY 279 ISSPQELGHAFIDNVRVPLELDQIIRLPFNFTSTPQETLSIRHFDLVE 328  
 Db 335 LSSISELKHLSGHAKVKPDPKTKCKOECLITTFQDVYFVSSEFEDA 384  
 RESULT 6  
 TX3H\_PHASP STANDARD; PRT; 491 AA.  
 AC P11982;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Tyrosine 3-monooxygenase (EC 1.14.16.2) (Tyrosine 3-hydroxylase) (TH).  
 GN TH.  
 OS Phasianidae sp. (Quail).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae.  
 OX NCBI\_TaxID=9006;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Adrenal gland;  
 RX MEDLINE=88089590; PubMed=2447231;  
 RA Fauguet M, Grima B, Lamouroux A, Mallet J;  
 RT "Cloning of quail tyrosine hydroxylase; amino acid homology with  
 RT other hydroxylases discloses functional domains.";  
 RL J. Neurochem. 50:142-148(1988).  
 CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE PHYSIOLOGY OF ADRENERGIC  
 CC NEURONES.  
 CC -!- CATALYTIC ACTIVITY: L-tyrosine + tetrahydropteridine + O(2) = 3,4-  
 CC dihydroxy-L-phenylalanine + dihydropteridine + H(2)O.  
 CC -!- COFACTOR: FERROUS ION.  
 CC -!- ENZYME REGULATION: PHOSPHORYLATION LEADS TO AN INCREASE IN THE  
 CC CATALYTIC ACTIVITY.  
 CC -!- PATHWAY: Catecholamine biosynthesis; first step.  
 CC -!- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID  
 CC HYDROXYLASES FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M24778; AAA49514.1; -;  
 CC PIR; A28582; A28582.  
 CC HSSP; P04177; ITOH.  
 CC InterPro; IPR001273; Aaa\_hydroxylase.  
 CC Pfam; PF00351; bipterin\_H; 1.  
 CC PRINTS; PR00372; FWHYDRXLASE.  
 CC TIGRFAMs; TIGR01269; Tyr\_3\_monoox; 1.  
 CC PROSITE; PS00367; BIOTERIN\_HYDROXYL; 1.  
 CC Catecholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;  
 CC Neurotransmitter biosynthesis; Phosphorylation.  
 CC MOD RES 40 40 PHOSPHORYLATION (BY PKA) (BY SIMILARITY).  
 CC DOMAIN 51 55 POLY-ALA.  
 CC METAL 324 324 IRON (BY SIMILARITY).  
 CC METAL 329 329 IRON (BY SIMILARITY).  
 CC METAL 369 369 IRON (BY SIMILARITY).  
 CC SEQUENCE 491 AA; 56066 MW; AFB363220F70C0A0 CRC64;  
 CC  
 CC Query Match 12.1%; Score 229.5; DB 1; Length 491;  
 CC Best Local Similarity 29.1%; Pred. No. 8.7e-11;  
 CC Matches 66; Conservative 34; Mismatches 100; Indels 27; Gaps 4;  
 CC  
 CC QY 107 WYRLSSRFLSKYCRPFLLDYLEAFGLSLDFLDH-----QAVIKFPELETHFSY 158  
 CC Db 226 WKEVSTLSKSLYPTHACK---EYLEAFNLKFCGYNNENNIPQLEEVSRFLKERTGFQLR 282

```

QY 159 PVSGFVAPHOYLSLQDRYFPIASVWRTLDKDNFSLTDLHLGHVWMLHPSESEFF 218
DB 283 FVRGLSARDPLASLAFRFVQCTOYRHHASSPMHSPEDPCCHLGHVPMADKTPAQS 342
QY 219 INMGRLFTKVIKVOALPSKKORIOTLQSNLIAIVRCFWFTVESGLIENHGRKAYGAVL 278
DB 343 QDIG-----LASIGATDEIEKTLATL-----WFTVEGGLCKRONGIYVAGAGL 386
QY 279 ISSPQELGAHFDINVRVLPLELDQIRLPENTSTPOETLFSIRHDE 325
DB 387 LSSYGEHLHSLSDPEVRDPDPAAVAVCPQDPQVYPVYFVSESPSD 433

RESULT 7
TRSH RAT STANDARD; PRT; 444 AA.
ID TRSH RAT STANDARD; PRT; 444 AA.
AC P09810;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrocinophan 5-monooxygenase (EC 1.14.16.4) (Tyrocinophan 5-hydroxylase).
GN Tyrocinophan 5-monooxygenase (EC 1.14.16.4) (Tyrocinophan 5-hydroxylase).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Pineal gland;
RX MEDLINE=88244702; PubMed=3379411;
RA Darmon M.C., Gilbert B., Levitel V., Ehret M., Maitre M., Mallet J.;
RT "Sequence of two mRNAs encoding active rat tyrocinophan hydroxylase.";
RL J. Neurochem. 51:312-316(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91245924; PubMed=1645430;
RA Kim K.S., Messel T.C., Stone D.M., Carver C.H., Joh T.H., Park D.H.;
RT "Molecular cloning and characterization of cDNA encoding tyrocinophan
hydroxylase from rat central serotonergic neurons.";
RL Brain Res. Mol. Brain Res. 9:277-283(1991).
RN [3]
RP SEQUENCE OF 167-261 FROM N.A.
RX MEDLINE=87005247; PubMed=2875901;
RA Darmon M.C., Grima B., Cash C.D., Maitre M., Mallet J.;
RT "Isolation of a rat pineal gland cDNA clone homologous to tyrosine
and phenylalanine hydroxylases.";
RL FEBS Lett. 206:43-46(1986).
RL FEBS Lett. 206:43-46(1986).
CC -1- CATALYTIC ACTIVITY: L-tyrocinophan + tetrahydropteridine + O(2) = 5-
hydroxy-L-tyrocinophan + dihydropteridine + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- PATHWAY: THIS IS THE RATE-LIMITING ENZYME IN THE BIOSYNTHESIS OF
SEROTONIN IN THE CENTRAL NERVOUS SYSTEM AND CATALYZES THE FIRST
STEP OF THE SYNTHESIS OF MELATONIN IN THE PINEAL GLAND.
CC -1- SUBUNIT: MULTIMER OF IDENTICAL SUBUNITS.
CC -1- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID
HYDROXYLASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M28000; AAA2262.1; -
DR EMBL; X53501; CAA37579.1; -
DR PIR; U10034; MWRW
DR PIR; A24367; A24367.
DR PIR; A60034; A60034.
DR HSP; P04177; I10H.
DR InterPro; IPR002912; ACT.

```

```

DR InterPro; IPR001273; Aaa_hydroxylase.
DR Pfam; PF00351; Biopterin_H; 1.
DR Pfam; PF01842; ACT; 1.
DR PRINTS; PRO0372; PWHYDRXLASE.
DR TIGRfams; TIGR01270; TYP_5_monoox; 1.
DR PROSITE; PS00367; BIOPTERIN_HYDROXYL; 1.
KW Oxidoreductase; Monooxygenase; Serotonin biosynthesis; Iron;
KW Phosphorylation.
FT MOD_RES 58 58 PHOSPHORYLATION (BY PRA) (POTENTIAL).
FT METAL 272 272 IRON (BY SIMILARITY).
FT METAL 277 277 IRON (BY SIMILARITY).
FT METAL 317 317 IRON (BY SIMILARITY).
SQ SEQUENCE 444 AA; 51068 MW; C3CF5245727C825 CRC64;

Query Match 12.0%; Score 227.5; DB 1; Length 444;
Best Local Similarity 27.4%; Pred. No. 1.1e-10;
Matches 63; Conservative 43; Mismatches 97; Indels 27; Gaps 4;

QY 107 WYRLSSRSFSLKWCYCPREFIDYLEAFGLSDPDH-----QAVIFPELEHNSFY 158
DB 174 WGITREMLAKLPFHACR---ELRNPULSKYIGRENVQLEVSNFLKERTGFSIR 230
QY 159 PVSGFVAPHOYLSLQDRYFPIASVWRTLDKDNFSLTDLHLGHVWMLHPSESEFF 218
DB 231 FVAGYLSPPDFLSGLAFRFVQCTOYRHHSSDPLTPTEPCHLGHVPLAEPSPAQS 290
QY 219 INMGRLFTKVIKVOALPSKKORIOTLQSNLIAIVRCFWFTVESGLIENHGRKAYGAVL 278
DB 291 QDIG-----LASIGATDEIEKTLATL-----WFTVEGGLCKRONGIYVAGAGL 334
QY 279 ISSPQELGAHFDINVRVLPLELDQIRLPENTSTPOETLFSIRHDE 325
DB 335 LSSISLRLHSLGSHAKVFPDPKVAACKQECILTSFDVYFVSESPSDAKE 384

RESULT 8
TRSH MOUSE STANDARD; PRT; 447 AA.
ID TRSH MOUSE STANDARD; PRT; 447 AA.
AC P15532;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrocinophan 5-monooxygenase (EC 1.14.16.4) (Tyrocinophan 5-hydroxylase).
GN Tyrocinophan 5-monooxygenase (EC 1.14.16.4) (Tyrocinophan 5-hydroxylase).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90243261; PubMed=2110547;
RA Scoll J., Kozak C.A., Goldman D.;
RT "Characterization and chromosomal mapping of a cDNA encoding
tyrocinophan hydroxylase from a mouse mastocytoma cell line.";
RL Genomics 7:88-96(1990).
RL Genomics 7:88-96(1990).
CC -1- CATALYTIC ACTIVITY: L-tyrocinophan + tetrahydropteridine + O(2) = 5-
hydroxy-L-tyrocinophan + dihydropteridine + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- PATHWAY: THIS IS THE RATE-LIMITING ENZYME IN THE BIOSYNTHESIS OF
SEROTONIN IN THE CENTRAL NERVOUS SYSTEM AND CATALYZES THE FIRST
STEP OF THE SYNTHESIS OF MELATONIN IN THE PINEAL GLAND.
CC -1- SUBUNIT: MULTIMER OF IDENTICAL SUBUNITS.
CC -1- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID
HYDROXYLASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; J04758; AAA63401.1; -.
DR PIR; A34582; A34582.
DR HSSP; P04177; ITOH.
DR MGI; MGI:98796; Tph.
DR InterPro; IPR002912; ACT.
DR Pfam; PF00351; bioterin_H; 1.
DR PRINTS; PR00372; FWHYDRXLASE.
DR TIGRfams; TIGR01270; Ttp 5 monoox. 1.
DR PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
DR Oxidoreductase; Monooxygenase; Serotonin biosynthesis; Iron;
KW Phosphorylation.
FT MOD_RES 61 61 PHOSPHORYLATION (BY PKA) (POTENTIAL).
FT METAL 275 275 IRON (BY SIMILARITY).
FT METAL 280 280 IRON (BY SIMILARITY).
FT METAL 320 320 IRON (BY SIMILARITY).
SQ SEQUENCE 447 AA; 51343 MW; 16C839F22A138BCA CRC64;

Query Match 12.0%; Score 226.5; DB 1; Length 447;
Best Local Similarity 27.4%; Pred. No. 1.3e-10;
Matches 63; Conservative 43; Mismatches 97; Indels 27; Gaps 4;

QY 107 WYRLSSRFSIMKSYCPFRFDLYEAFGLLSDFLDH-----QAVIKFPELETHFSY 158
DB 177 WGTIFRELNKLYPTHACR---EYLRLNPLLSKYCGYREDNIPQLEDVSNFLKERTGFSIR 233
QY 159 PVSGFVAPHOVLSDQRYFPPIASVMRTLDKNFSLTPDLIHDLLGHVPMWLLHPSSEFF 218
DB 234 PVAGYLSPRDFLSGLAFRVFCHCTQYVVRHSSDPDYTPEDTCHELLGHVPLLAESPFAQFS 293
QY 219 INMGRLFTKVIKQVALPSKKORIOTLOSNIATVRCFWTVESGLIENHEGRKAYGAVL 278
DB 294 QEIG-----LASLGASEEAVOKLAT-----CYFFTVFGLCKQDQQLRVFGAGL 337
QY 279 ISSQELCHAFIDNVRVLPLELDQIIRLPFWNTSTPQETLPSIRHFDLVE 328
DB 338 LSSISELKHLSGHAKVKPPDKPIACKQELITSFQDYVFSSEFDEAKE 387

RESULT 9
TRSH RABIT STANDARD; PRT; 444 AA.
AC P17290; Q29523;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tryptophan 5-monoxygenase (EC 1.14.16.4) (Tryptophan 5-hydroxylase).
GN TPH.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87289639; PubMed=3475690;
RA Grenett H.E., Jedley F.D., Reed L.L., Woo S.L.C.;
RT Full-length cDNA for rabbit tryptophan hydroxylase: functional
RL domains and evolution of aromatic amino acid hydroxylases.;
RN Proc. Natl. Acad. Sci. U.S.A. 84:5530-5534 (1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95077422; PubMed=7986090;
RA Tipper J.P., Citron B.A., Ribeiro P., Kaufman S.;
RT "Cloning and expression of rabbit and human brain tryptophan
RL hydroxylase cDNA in Escherichia coli.";
RN Arch. Biochem. Biophys. 315:445-453 (1994).
CC -!- CATALYTIC ACTIVITY: L-tryptophan + tetrahydropteridine + O(2) = 5-
CC hydroxy-L-tryptophan + dihydropteridine + H(2)O.
CC -!- COPACITOR; FERROUS ION.
CC -!- PATHWAY: THIS IS THE RATE-LIMITING ENZYME IN THE BIOSYNTHESIS OF
CC SEROTONIN IN THE CENTRAL NERVOUS SYSTEM AND CATALYZES THE FIRST

```

---

```

CC STEP OF THE SYNTHESIS OF MELATONIN IN THE PINEAL GLAND.
CC -!- SUBUNIT: MULTIMER OF IDENTICAL SUBUNITS.
CC -!- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M17250; AAA31487.1; -.
CC PIR; A32699; AAA67051.1; -.
CC HSSP; P04177; ITOH.
CC InterPro; IPR002912; ACT.
CC InterPro; IPR001273; Aaa hydroxylase.
CC Pfam; PF00351; bioterin_H; 1.
CC PRINTS; PR00372; FWHYDRXLASE.
CC TIGRfams; TIGR01270; Ttp 5 monoox; 1.
CC PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
CC Oxidoreductase; Monooxygenase; Serotonin biosynthesis; Iron;
KW Phosphorylation.
FT MOD_RES 58 58 PHOSPHORYLATION (BY PKA) (POTENTIAL).
FT METAL 272 272 IRON (BY SIMILARITY).
FT METAL 277 277 IRON (BY SIMILARITY).
FT METAL 317 317 IRON (BY SIMILARITY).
FT CONFLICT 102 102 M -> L (IN REF. 1).
FT CONFLICT 151 151 L -> S (IN REF. 2).
FT CONFLICT 202 203 KY -> ND (IN REF. 1).
FT CONFLICT 207 207 R -> Q (IN REF. 2).
FT CONFLICT 390 390 T -> K (IN REF. 1).
SQ SEQUENCE 444 AA; 51118 MW; BF182451B28ECD80 CRC64;

Query Match 11.9%; Score 225.5; DB 1; Length 444;
Best Local Similarity 27.0%; Pred. No. 1.6e-10;
Matches 62; Conservative 44; Mismatches 97; Indels 27; Gaps 4;

QY 107 WYRLSSRFSIMKSYCPFRFDLYEAFGLLSDFLDH-----QAVIKFPELETHFSY 158
DB 174 WGTIFRELNKLYPTHACR---EYLKMLPLLSKYCGYREDNIPQLEDISNFKERTGFSIR 230
QY 159 PVSGFVAPHOVLSDQRYFPPIASVMRTLDKNFSLTPDLIHDLLGHVPMWLLHPSSEFF 218
DB 231 PVAGYLSPRDFLSGLAFRVFCHCTQYVVRHSSDPDYTPEDTCHELLGHVPLLAESPFAQFS 290
QY 219 INMGRLFTKVIKQVALPSKKORIOTLOSNIATVRCFWTVESGLIENHEGRKAYGAVL 278
DB 291 QEIG-----LASLGASEEAVOKLAT-----CYFFTVFGLCKQDQQLRVFGAGL 334
QY 279 ISSQELCHAFIDNVRVLPLELDQIIRLPFWNTSTPQETLPSIRHFDLVE 328
DB 335 LSSISELKHLSGHAKVKPPDKPIYKQELITTFQDYVFSSEFDEAKE 384

RESULT 10
TY3H ANGAN STANDARD; PRT; 488 AA.
ID TY3H ANGAN
AC O42091;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine 3-monoxygenase (EC 1.14.16.2) (Tyrosine 3-hydroxylase) (TH).
GN TH.
OS Anguilla anguilla (European freshwater eel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Anguilliformes; Anguillidae;
OC Anguilla
OX NCBI_TaxID=7936;
RN [1]

```

```

RE SEQUENCE FROM N.A.
RX MEDLINE=98344760; PubMed=9681435;
RA Bouland S., Biguet N.F., Vidal B., Veron M., Mallet J.,
RT Vincent J.D., Ducour S., Vernier P.,
RT "Tyrosine hydroxylase in the european eel (Anguilla anguilla): cDNA
RT Cloning, brain distribution, and phylogenetic analysis.";
RL J. Neurochem. 71:460-470(1998).
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE PHYSIOLOGY OF ADRENERGIC
CC NEURONES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: L-tyrosine + tetrahydropteridine + O(2) = 3,4-
CC dihydroxy-L-phenylalanine + dihydropteridine + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- ENZYME REGULATION: PHOSPHORYLATION LEADS TO AN INCREASE IN THE
CC CATALYTIC ACTIVITY (BY SIMILARITY).
CC -1- PATHWAY: Catecholamine biosynthesis; first step.
CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ000731; CA04264.1; -.
CC HSP; P04177; ITOH.
CC InterPro: IPR001273; Aaa_hydroxylase.
CC Pfam; PF00351; bioprotein_H; 1.
CC PRINTS; PR00372; FYMHYDRLXASE.
CC TIGRFAMs; TIGR01269; Tyr_3_monoox; 1.
CC PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
CC KX Neurotransmitter biosynthesis; Oxidoreductase; Monoxygenase; Iron;
CC MOD_RES 38 38 PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
CC FT METAL 321 321 IRON (BY SIMILARITY).
CC FT METAL 326 326 IRON (BY SIMILARITY).
CC FT METAL 366 366 IRON (BY SIMILARITY).
CC SQ SEQUENCE 488 AA; 55490 MW; 575353BA39EBD448 CRC64;
CC -----
Query Match 11.9%; Score 225.5; DB 1; Length 488;
Best Local Similarity 28.1%; Pred. No. 1.8e-10;
Matches 65; Conservative 38; Mismatches 99; Indels 29; Gaps 5;
CC -----
OY 107 MYRLSSRFLMKSY-CRPFLLDYLAFLGSLDFLDH-----QAVIKFELETHFSY 157
DB 223 WREYVSTLRDLTYTHACS-----EHLERFLRLRHGCGYSPNSIPQLEVDVSHLKERITGFL 278
OY 158 YPVSGFVAHPHOYSLDODRFPIASVWRTLDGNSLTPDILHDLGHVPLMLHPSFSEF 217
DB 279 RPYVGLLSARDPLASLAFRFQCTGYIRIASSPMHSPBDCHEHLGHVPLADRTAQS 338
OY 218 FINNGRLFTKVIENQALPSKQRIOTLOSNIJAVRCFWFTVSGLIENHGRKAYGAV 277
DB 339 SQNG-----LASLGASEEDIKST-----LYWTFVEGLCKQGDVYKAYGAG 382
OY 278 LISSPOELGAFIDNVRVLPLELDQIIRLPFTSTPQETLSIRHFDLVE 328
DB 383 LLSSYGEIVHSLSDPERREFDEAAAPYODVQSYVFSSESFTDAKE 433
CC -----
RESULT 11
TY3H MOUSE STANDARD; PRT; 498 AA.
AC P24529;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine 3-monooxygenase (EC 1.14.16.2) (Tyrosine 3-hydroxylase) (TH).
GN TH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RX SEQUENCE FROM N.A.
RA MEDLINE=91248263; PubMed=1674869;
RA Ichikawa S., Sasaka T., Nagatsu T.;
RT "Primary structure of mouse tyrosine hydroxylase deduced from its
RT cDNA.";
RL Biochem. Biophys. Res. Commun. 176:1610-1616(1991).
CC [2]
CC SEQUENCE OF 1-30 FROM N.A.
CC STRAIN=BALB/c;
RA Morgan W.W., Bernudez J., Sharp Z.D.;
RL Submitted (Apr-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE PHYSIOLOGY OF ADRENERGIC
CC NEURONES.
CC -1- CATALYTIC ACTIVITY: L-tyrosine + tetrahydropteridine + O(2) = 3,4-
CC dihydroxy-L-phenylalanine + dihydropteridine + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- ENZYME REGULATION: PHOSPHORYLATION LEADS TO AN INCREASE IN THE
CC CATALYTIC ACTIVITY.
CC -1- PATHWAY: Catecholamine biosynthesis; first step.
CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M69200; AAA0434.1; -.
CC DR PIR; J00068; JN0068.
CC DR HSP; P04177; ITOH.
CC MGD; MG1:98735; Th.
CC InterPro: IPR001273; Aaa_hydroxylase.
CC Pfam; PF00351; bioprotein_H; 1.
CC PRINTS; PR00372; FYMHYDRLXASE.
CC TIGRFAMs; TIGR01269; Tyr_3_monoox; 1.
CC PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
CC KX Neurotransmitter biosynthesis; Oxidoreductase; Monoxygenase; Iron;
CC MOD_RES 19 19 PHOSPHORYLATION (BY CMK2) (BY
CC SIMILARITY)
CC FT METAL 31 31 PHOSPHORYLATION (BY SIMILARITY).
CC FT MOD_RES 40 40 POLY-ALA.
CC FT DOMAIN 51 59
CC FT METAL 331 331 IRON (BY SIMILARITY).
CC FT METAL 336 336 IRON (BY SIMILARITY).
CC FT METAL 376 376 IRON (BY SIMILARITY).
CC SQ SEQUENCE 498 AA; 55992 MW; 62790179664F6DC6 CRC64;
CC -----
Query Match 11.9%; Score 225.5; DB 1; Length 498;
Best Local Similarity 27.5%; Pred. No. 1.8e-10;
Matches 66; Conservative 42; Mismatches 101; Indels 31; Gaps 5;
CC -----
OY 107 MYRLSSRFLMKSY-CRPFLLDYLAFLGSLDFLDH-----QAVIKFELETHFSY 158
DB 223 WREYVSTLRDLTYTHACS-----EHLERFLRLRHGCGYSPNSIPQLEVDVSHLKERITGFL 278
OY 159 YPVSGFVAHPHOYSLDODRFPIASVWRTLDGNSLTPDILHDLGHVPLMLHPSFSEF 218
DB 290 RPYVGLLSARDPLASLAFRFQCTGYIRIASSPMHSPBDCHEHLGHVPLADRTAQS 349
OY 219 FINNGRLFTKVIENQALPSKQRIOTLOSNIJAVRCFWFTVSGLIENHGRKAYGAV 278
DB 350 QDID-----LASLGASEDEIKST-----YVWTFVEGLCKQNGELKAYGAG 393
OY 279 ISSPOELGAFIDNVRVLPLELDQIIRLPFTSTPQETLSIRHFDLVE 328
DB 383 LLSSYGEIVHSLSDPERREFDEAAAPYODVQSYVFSSESFTDAKE 433
CC -----

```

Db 394 LSSYCELLHSLSEBEVRAFPDPTAAVQPDQTYQVYFVSEBSFSDAKDLNVAIRIQ 453

## RESULT 12

TRSH\_XENLA STANDARD; PRT; 481 AA.  
ID Q92142;  
AC 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Tryptophan 5-monooxygenase (EC 1.14.16.4) (tryptophan 5-hydroxylase).  
GN TPH.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Retina;  
RX MEDLINE=94246419; PubMed=8189245;  
RA Green C.B., Besharse J.C.;  
RT "Tryptophan hydroxylase expression is regulated by a circadian clock  
in Xenopus laevis retina.";  
RL J. Neurochem. 62:2420-2428(1994).  
CC -|- CATALYTIC ACTIVITY: L-tryptophan + tetrahydropteridine + O(2) = 5-  
hydroxy-L-tryptophan + dihydropteridine + H(2)O.  
CC -|- COPACITOR: FERROUS ION.  
CC -|- PATHWAY: THIS IS THE RATE-LIMITING ENZYME IN THE BIOSYNTHESIS OF  
SEROTONIN IN THE CENTRAL NERVOUS SYSTEM AND CATALYZES THE FIRST  
STEP OF THE SYNTHESIS OF MELATONIN IN THE PINEAL GLAND.  
CC -|- SUBUNIT: MULTIMER OF IDENTICAL SUBUNITS (BY SIMILARITY).  
CC -|- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID  
HYDROXYLASES FAMILY.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; L20679; AAA21306.1; --  
DR HSSP; P04177; ITOH.  
DR InterPro; IPR002912; ACT.  
DR InterPro; IPR001273; Aaa hydroxylase.  
DR Pfam; PF00351; bioterin\_H; 1.  
DR Pfam; PF01842; ACT; 1.  
DR PRINTS; PR00372; FWHYDRXLASE.  
DR TIGRFS; TIGR01270; TYP\_5\_MONOXYL; 1.  
DR PROSITE; PS00367; BIOTERIN\_HYDROXYL; 1.  
KW Oxidoreductase; Monooxygenase; Serotonin biosynthesis; Iron.  
FT METAL 309 309 IRON (BY SIMILARITY).  
FT METAL 314 314 IRON (BY SIMILARITY).  
FT METAL 354 354 IRON (BY SIMILARITY).  
SQ SEQUENCE 481 AA; 55406 MW; 6852C33EFD8DEBA0 CRC64;

Query Match 11.8%; Score 223.5; DB 1; Length 481;  
Best Local Similarity 29.5%; Pred. No. 2.5e-10;  
Matches 70; Conservative 38; Mismatches 90; Indels 39; Gaps 9;

QY 107 WYRLSSRFLSKSYCPFRFLDYLAFLGLSLDFLDH-----QAVIKPELETHFSY 158  
Db 211 WGTFRRLNKLPHYTHACR---EYLNKLPLSKHCKYREDNIPQLEDVSRFLRBTGFTIR 267  
QY 159 PVSGFVAPHVYLSLDQRPPIASVNRITLDKNFSL-TPDLIHLLGHVPEWLLHPFSFSE 217  
Db 268 PVAGYLSRDLFLAGFRVCHYQVYR-DSPLNTPEPTDCHELLGHVLLAEPSPFAQ 326  
QY 218 FINMGFLTKVIEKVALPSKQRIQTLQSLNLAIVRCFWFVSGLIENHGRKAYGAV 277  
Db 327 SQETG-----LASLGASDEAVQKLAT-----CYFFTFVFLCKQEGKLKVGAG 370

QY 278 LISSPOELGHAFIDNVRVLPLELDQIIRLPNTSTPQRTLFSIRHFDLVELTSKLE 334  
Db 371 LLSSISELKHSLSGNAKVKFPD-----PWTCTN-QECI--ITSFOELVFSSEPE 417

## RESULT 13

TY3H\_RAT STANDARD; PRT; 498 AA.  
ID P04177;  
AC 20-MAR-1987 (Rel. 04, Created)  
DT 20-MAR-1987 (Rel. 04, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Tyrosine 3-monooxygenase (EC 1.14.16.2) (tyrosine 3-hydroxylase) (TH).  
GN TH.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85113249; PubMed=2857492;  
RA Grima B., Lamouroux A., Blanot F., Faucon Biquet N., Mallet J.;  
RT "Complete coding sequence of rat tyrosine hydroxylase mRNA.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:617-621(1985).  
CC [2]  
RP SEQUENCE FROM N.A.  
RA Anton X.X., Manaster J.S., Kordower X.X., Markham X.X., Bredeesen D.E.;  
RN Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.  
CC [3]  
RP PHOSPHORYLATION  
RX MEDLINE=91170235; PubMed=1672315;  
RA Haycock J.W., Haycock D.A.;  
RT "Tyrosine hydroxylase in rat brain dopaminergic nerve terminals.  
Multiple-site phosphorylation in vivo and in synaptosomes.";  
RL J. Biol. Chem. 266:5650-5657(1991).  
CC [4]

X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 164-498.  
EMBL=97372896; PubMed=9228951;  
RA Goodwill K.E., Sabatier C., Marks C., Raag R., Fitzpatrick P.F.,  
RA Stevens R.C.;  
RT "Crystal structure of tyrosine hydroxylase at 2.3 A and its  
implications for inherited neurodegenerative diseases.";  
RL Nat. Struct. Biol. 4:578-585(1997).

CC -|- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE PHYSIOLOGY OF ADRENERGIC  
NEURONES.

CC -|- CATALYTIC ACTIVITY: L-tyrosine + tetrahydropteridine + O(2) = 3,4-  
dihydroxy-L-phenylalanine + dihydropteridine + H(2)O.

CC -|- COPACITOR: FERROUS ION.

CC -|- ENZYME REGULATION: PHOSPHORYLATION LEADS TO AN INCREASE IN THE  
CATALYTIC ACTIVITY.

CC -|- PATHWAY: Catecholamine biosynthesis; first step.

CC -|- SUBUNIT: HOMOTETRAMER.

CC -|- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID  
HYDROXYLASES FAMILY.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; M10244; AAA42257.1; --  
DR EMBL; L22651; AAA42258.1; --  
DR PIR; A00510; WHRTY.  
DR PDB; 1TOH; 03-JUN-98.  
DR InterPro; IPR001273; Aaa hydroxylase.  
DR Pfam; PF00351; bioterin\_H; 1.  
DR PRINTS; PR00372; FWHYDRXLASE.  
DR TIGRFS; TIGR01269; TYX\_3\_MONOXYL; 1.  
DR PROSITE; PS00367; BIOTERIN\_HYDROXYL; 1.

KW Catecholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;  
 KM Neurotransmitter biosynthesis; Phosphorylation (by CAMK2).  
 FT MOD\_RES 19 19 PHOSPHORYLATION (BY CAMK2).  
 FT MOD\_RES 31 31 PHOSPHORYLATION (BY PKA).  
 FT MOD\_RES 40 40 PHOSPHORYLATION (BY PKA).  
 FT DOKA1N 51 59 POLY-ALA.  
 FT METAL 331 331 IRON.  
 FT METAL 336 336 IRON.  
 FT METAL 376 376 IRON.  
 SQ SEQUENCE 498 AA; 55965 MW; 17F7E003D29218C5 CRC64;

Query Match 11.8%; Score 223; DB 1; Length 498;  
 Best Local Similarity 26.7%; Pred. No. 2.9e-10;  
 Matches 65; Conservative 39; Mismatches 103; Indels 36; Gaps 5;

QY 112 SSRSFKMS-----YCRPFLDYLEAFGLSDPLDH-----QAVIKPELETHF 155  
 DB 227 ABEIATNKEVYVTLKGIYATHACREHLEGFQLEKRCYGRDSITPLEDVSRLKERTGF 286  
 QY 156 STYPSGFVAHQYLSLDQRYPPIASVWRTLDKDNFSLTPDLIHDLGHVPMILHPSFS 215  
 DB 287 QLRVAGLLSRDPLASLAFRVQCTQYIRHASSPMHSPEDPCHELLGHVPMILADRTFA 346  
 QY 216 EFTINMGELFTKVIKYOALPSKKORIQTLOSLNLAIVRCFMFTVESGLIENHGRKAYG 275  
 DB 347 QFSODIG-----LASIGASDEIEKLTST-----VWFVTEFGLCKONGELKAYG 330  
 QY 276 AVLLSPQELGAFIDNVRVLPLELDQIIRLPFNTSTPQETLFSIRHF-----DELVELTS 331  
 DB 391 AGLSSVGEHLHSLSEPERVAPDPTAAVQPYODQYQVYFVSESFNDAKDLRYNYS 450  
 QY 332 KLE 334  
 DB 451 RIQ 453

RESULT 14  
 TY3H BOVIN STANDARD; PRT; 490 AA.  
 AC p12789;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Tyrosine 3-monooxygenase (EC 1.14.16.2) (Tyrosine 3-hydroxylase) (TH).  
 GN Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 NC NCB1\_TaxID=9913;  
 RP MEDLINE=88259287; PubMed=2898537;  
 RA D'Mello S.R., Weisberg E.P., Stachowiak M.K., Turzai L.M., Glozio A.E.,  
 RA Kaplan B.B.;  
 RT "Isolation and nucleotide sequence of a cDNA clone encoding bovine  
 RT adrenal tyrosine hydroxylase: comparative analysis of tyrosine  
 RT hydroxylase gene products.";  
 RL J. Neurosci. Res. 19:440-449 (1988).  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88274405; PubMed=2899135;  
 RA Saadat S., Stehle A.D., Lamouroux A., Mallet J., Thoenen H.;  
 RT "Predicted amino acid sequence of bovine tyrosine hydroxylase and its  
 RT similarity to tyrosine hydroxylases from other species.";  
 RL J. Neurochem. 51:572-578 (1988).  
 RP SEQUENCE OF 153-169.  
 RX MEDLINE=88183482; PubMed=2895648;  
 RA Abate C., Smith J.A., Jon T.H.;  
 RT "Characterization of the catalytic domain of bovine adrenal tyrosine  
 RT hydroxylase.";  
 RL Biochem. Biophys. Res. Commun. 151:1446-1453 (1988).

RN [4]  
 RP SEQUENCE OF 1-27.  
 RC TISSUE=Adrenal medulla;  
 RX MEDLINE=88163736; PubMed=2894860;  
 RA Haavik J., Andersson K.K., Petersen L., Flatmark T.;  
 RT "Soluble tyrosine hydroxylase (tyrosine 3-monooxygenase) from bovine  
 RT adrenal medulla: large-scale purification and physicochemical  
 RT properties.";  
 RL Biochim. Biophys. Acta 953:142-156 (1988).  
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE PHYSIOLOGY OF ADRENERGIC  
 CC NEURONES.  
 CC -1- CATALYTIC ACTIVITY: L-tyrosine + tetrahydropteridine + O(2) = 3,4-  
 CC dihydroxy-L-phenylalanine + dihydropteridine + H(2)O.  
 CC -1- COFACTOR: FERROUS ION.  
 CC -1- ENZYME REGULATION: PHOSPHORYLATION LEADS TO AN INCREASE IN THE  
 CC CATALYTIC ACTIVITY.  
 CC -1- PATHWAY: catecholamine biosynthesis; first step.  
 CC -1- SIMILARITY: BELONGS TO THE BIOTRAN-DEPENDENT AROMATIC AMINO ACID  
 CC HYDROXYLASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M36794; AAA30779.1; -;  
 DR EMBL; M36705; AAA30798.1; -;  
 DR PIR; A27687; A27687.  
 DR PIR; J10039; J10039.  
 DR HSSP; P04177; ITOH.  
 DR InterPro: IPR001273; Aaa hydroxylase.  
 DR Pfam: PF00351; bioprotein H; 1.  
 DR PRINTS: PR00372; FWHYHYDXLASE.  
 DR TIGRfam: TIGR01269; TYR 3-monoox; 1.  
 DR PROSITE: PS00367; BIOTRAN HYDROXYL.  
 DR Catecholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;  
 KM Neurotransmitter biosynthesis; Phosphorylation.  
 FT INITIAT 0  
 FT MOD\_RES 39 39 PHOSPHORYLATION (BY PKA).  
 FT METAL 323 323 IRON (BY SIMILARITY).  
 FT METAL 328 328 IRON (BY SIMILARITY).  
 FT METAL 368 368 IRON (BY SIMILARITY).  
 FT CONFLICT 64 67 AAWL -> GSLV (IN REF. 2).  
 FT CONFLICT 72 72 E -> K (IN REF. 2).  
 FT CONFLICT 82 82 P -> R (IN REF. 2).  
 FT CONFLICT 85 85 R -> V (IN REF. 2).  
 FT CONFLICT 283 283 A -> V (IN REF. 2).  
 FT CONFLICT 320 320 E -> D (IN REF. 2).  
 FT CONFLICT 327 329 GHV -> AHG (IN REF. 2).  
 FT CONFLICT 379 379 K -> R (IN REF. 2).  
 FT CONFLICT 470 470 H -> R (IN REF. 2).  
 SQ SEQUENCE 490 AA; 54992 MW; 5F54PF233CEBAED CRC64;

Query Match 11.7%; Score 221.5; DB 1; Length 490;  
 Best Local Similarity 27.1%; Pred. No. 3.7e-10;  
 Matches 65; Conservative 42; Mismatches 102; Indels 31; Gaps 5;

QY 107 WYRLSSRSFKMSYCRPFLDYLEAFGLSDPLDH-----QAVIKPELETHFSTY 158  
 DB 225 WKVEYTLKGLYTHACR-----EHLAFELRFGYEDRIPOLEDVSRLKERTGFQUR 281  
 QY 159 PVSGFVAHQYLSLDQRYPPIASVWRTLDKDNFSLTPDLIHDLGHVPMILHPSFSFEF 218  
 DB 282 PAAGLLSARDPLASLAFRVQCTQYIRHASSPMHSPEDPCHELLGHVPMILADRTFAQFS 341  
 QY 219 INNGRLFTKVIKYOALPSKKORIQTLOSLNLAIVRCFMFTVESGLIENHGRKAYGAVL 278  
 DB 342 QDQIG-----LASIGASDEIEKLTST-----LYMFTVEFGLCKONGELKAYGAGL 385  
 QY 279 ISSPQELGAFIDNVRVLPLELDQIIRLPFNTSTPQETLFSIRHF-----DELVELTSKLE 334





```

CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE PHYSIOLOGY OF ADRENERGIC
CC NEURONES.
CC -1- CATALYTIC ACTIVITY: L-tyrosine + tetrahydropteridine + O(2) = 3,4-
CC dihydroxy-L-phenylalanine + dihydropteridine + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- ENZYME REGULATION: PHOSPHORYLATION LEADS TO AN INCREASE IN THE
CC CATALYTIC ACTIVITY.
CC -1- PATHWAY: catecholamine biosynthesis; first step.
CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1, 2, 3 (SHOWN HERE) AND 4;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y00414; CAA6472.1; ALT_SEQ.
DR EMBL: M17588; AAA61179.1; -.
DR EMBL: X05290; CAA28908.1; -.
DR EMBL: M24791; AAA61173.1; -.
DR EMBL: M24787; AAA61173.1; JOINED.
DR EMBL: M24789; AAA61173.1; JOINED.
DR EMBL: M24791; AAA61170.1; JOINED.
DR EMBL: M24787; AAA61170.1; JOINED.
DR EMBL: M20911; AAA61167.1; -.
DR PIR: A27791; WHH011.
DR PIR: A27791; WHH012.
DR PIR: C27791; WHH013.
DR PIR: A26825; WHH014.
DR PIR: JE0012; JE0012.
DR PIR: JE0013; JE0013.
DR PIR: JE0014; JE0014.
DR HSSP: P04177; ITOH.
DR GeneW: HGNC:11782; TH.
DR MIM: 191290; -.
DR InterPro: IPR001273; Aaa_hydroxylase.
DR Pfam: PF00351; bioplerin_H; 1.
DR PRINTS: PR00372; FYMHYDRXLASE.
DR TIGRfams: TIGR01269; Tyr_3_monoox; 1.
DR PROSITE: PS00367; BIOPTERIN_HYDROXYL; 1.
KW catecholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;
KW alternative splicing; Neurotransmitter biosynthesis; Phosphorylation.
KW MOD RES 71 71 PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
FT DOMAIN 85 90 POLY-ALA.
FT METAL 361 361 IRON (BY SIMILARITY).
FT METAL 366 366 IRON (BY SIMILARITY).
FT METAL 406 406 IRON (BY SIMILARITY).
FT VARSPIC 1 33 MISSING (IN ISOFORM 4).
FT VARSPIC 34 34 O -> M (IN ISOFORM 4).
FT VARSPIC 31 61 MISSING (IN ISOFORM 1).
FT VARSPIC 35 61 MISSING (IN ISOFORM 2).
SQ SEQUENCE 528 AA; 58524 MW; B3DC34955A21074 CRC64;

Query Match 11.6%; Score 220; DB 1; Length 528;
Best Local Similarity 26.7%; Pred. No. 5, 3e-10;
Matches 65; Conservative 38; Mismatches 104; Indels 36; Gaps 5;

```

```

QY 276 AVLISSPDEIGAHFDNVRVLELEIDQIRLPNTSTPQETLSIRNF-----DELVELTS 331
DB 421 AGLLSYGELHCLSEPEIRAFDEAAVOPDQTYGVSVFSSPSDAKDKRSVNS 480
QY 332 KLE 334
DB 481 RIQ 483

RESULT 17
ID TY3H DROME STANDARD; PRT; 508 AA.
AC P18459;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Tyrosine 3-monooxygenase (EC 1.14.16.2) (Tyrosine 3-hydroxylase) (TH
DE (Protein Pale).
GN TH OR PLE.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Preygora; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN 11]
RP SEQUENCE FROM N.A.
RX MEDLINE=90165583; PubMed=2483109;
RA Neckmeyer W.S., Quinn W.G.,
RT "Isolation and characterization of the gene for Drosophila tyrosine
RT hydroxylase".
RU Neuron 21:167-1175(1989).
RL -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE PHYSIOLOGY OF ADRENERGIC
CC NEURONES.
CC -1- CATALYTIC ACTIVITY: L-tyrosine + tetrahydropteridine + O(2) = 3,4-
CC dihydroxy-L-phenylalanine + dihydropteridine + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- PATHWAY: catecholamine biosynthesis; first step.
CC -1- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U14395; AAA62877.1; -.
DR EMBL: X76209; CAA53802.1; -.
DR PIR: JN0010; JN0010.
DR HSSP: P04177; ITOH.
DR FLYBase: FBgn0005626; ple.
DR InterPro: IPR001273; Aaa_hydroxylase.
DR Pfam: PF00351; bioplerin_H; 1.
DR PRINTS: PR00372; FYMHYDRXLASE.
DR TIGRfams: TIGR01269; Tyr_3_monoox; 1.
DR PROSITE: PS00367; BIOPTERIN_HYDROXYL; 1.
KW catecholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;
KW Neurotransmitter biosynthesis.
FT METAL 338 338 IRON (BY SIMILARITY).
FT METAL 343 343 IRON (BY SIMILARITY).
FT METAL 383 383 IRON (BY SIMILARITY).
SQ SEQUENCE 508 AA; 57861 MW; 2019E0B0845E66F CRC64;

Query Match 11.6%; Score 219; DB 1; Length 508;
Best Local Similarity 28.6%; Pred. No. 6, 1e-10;
Matches 74; Conservative 37; Mismatches 94; Indels 54; Gaps 8;

```



```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M12337; AAA41843.1; -
DR EMBL: K02599; AAA41794.1; -
DR PIR: A00509; WHRTF.
DR PIR: A14970; A14970.
DR PIR: A25321; A25321.
DR PDB: 1PHZ; 30-APR-99.
DR PDB: 2PHM; 30-APR-99.
DR InterPro: IPR002912; Aaa_hydroxylase.
DR InterPro: IPR001273; Aaa_hydroxylase.
DR Pfam: PF01842; ACT.
DR Pfam: PF01842; ACT.
DR PRINTS: PR00372; FWHYDRYLASE.
DR TIGRfam: TIGR01268; PheHydrox_tetr; 1.
DR PROSITE: PS00367; BIOPTERIN_HYDROXYL; 1.
KW Oxidoreductase; Monooxygenase; Phosphorylation;
KW Phenylalanine catabolism; Iron; 3D-structure.
FT MOD_RES 16 16 PHOSPHORYLATION (BY PKA).
FT METAL 285 285 IRON.
FT METAL 290 290 IRON.
FT METAL 330 330 IRON.
SQ SEQUENCE 453 AA; 51821 MW; 365D9E8A7E498D52 CRC64;

Query Match 11.3%; Score 214; DB 1; Length 453;
Best Local Similarity 28.4%; Pred. No. 1.3e-09;
Matches 61; Conservative 37; Mismatches 89; Indels 28; Gaps 4;

OY 115 FSLWKSYCPHFLDYLEAFGLSDPLDH-QAVIKFPELEHPSYYPGAFAPHOYSL 173
DB 210 FPLEKCY-----GFRDNI PQLEDVSGFLQCTCFRLRPVAGLSSMDPLGGL 258
OY 174 QDRYFPIASVMTLDNDNSTLPDLDLHDLGHVPLHLPFSEPFIMNGRLFTVIEKNG 233
DB 259 AFRVHCTQYIRHGSKRMVTPEDDICHLLGHVPLFSDRSRQSGRTG-----LA 309
OY 234 ALPSKKQRIQTLOSNIAYRCWFVFSGLIENHGKRAYCAVLSSPQELGAFIDNV 293
DB 310 SLGAPDEYIEKLAT-----IYFVBEGLCKEGDSIRAYGAGLISFGLQYCUSDKP 362
OY 294 RVLPLELDQIIRLPENTSTPQETLFSIRHDELVE 328
DB 363 KLPLLEKTAQCEYSVTEFQPLVVAESFSDAKE 397

RESULT 20
PH4H HUMAN STANDARD; PRT; 452 AA.
AC P00439; Q16717;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phenylalanine-4-hydroxylase (EC 1.14.16.1) (PAH) (Phe-4-
DE monooxygenase).
GN PAH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RA [1]
RE SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA KMOX S.C.M.; Ledley F.D.; Dilella A.G.; Robson K.J.H.; Woo S.L.C.;
RT "Nucleotide sequence of a full-length complementary DNA clone and
RT amino acid sequence of human phenylalanine hydroxylase.";
```

```
RL Biochemistry 24:556-561(1985).
RL [2]
RL SEQUENCE FROM N.A.
RL Scriber C.R., Nowacki P.M., Byck S., Prevost L.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RL [3]
RL X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 117-424.
RL MEDLINE=98069646; PubMed=9406548;
RL Eriandesen H., Fusetti F., Martinez A., Hough E., Flatmark T.,
RL Stevens R.C.;
RL "Crystal structure of the catalytic domain of human phenylalanine
RL hydroxylase reveals the structural basis for phenylketonuria.";
RL Nat. Struct. Biol. 4:995-1000(1997).
RL [4]
RL X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 117-424.
RL MEDLINE=99060040; PubMed=9843368;
RL Eriandesen H., Flatmark T., Stevens R.C., Hough E.;
RL "Crystallographic analysis of the human phenylalanine hydroxylase
RL catalytic domain with bound catechol inhibitors at 2.0-A resolution.";
RL Biochemistry 37:15638-15646(1998).
RL [5]
RL X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF 117-452.
RL MEDLINE=98307935; PubMed=9642259;
RL Fusetti F., Eriandesen H., Flatmark T., Stevens R.C.;
RL "Structure of tetrameric human phenylalanine hydroxylase and its
RL implications for phenylketonuria.";
RL J. Biol. Chem. 273:16962-16967(1998).
RL [6]
RL REVIEW ON PKU VARIANTS.
RL MEDLINE=91348681; PubMed=1679029;
RL Konecki D.S., Lichter-Konecki U.;
RL "The phenylketonuria locus: current knowledge about alleles and
RL mutations of the phenylalanine hydroxylase gene in various
RL populations.";
RL Hum. Genet. 87:377-388(1991).
RL [7]
RL REVIEW ON PKU VARIANTS.
RL MEDLINE=91061429; PubMed=2246858;
RL Cotton R.G.;
RL "Heterogeneity of phenylketonuria at the clinical, protein and DNA
RL levels.";
RL J. Inher. Metab. Dis. 13:739-750(1990).
RL [8]
RL REVIEW ON PKU VARIANTS.
RL MEDLINE=93244626; PubMed=1301187;
RL Eriandesen H., Woo S.L.C.;
RL "Molecular basis of phenylketonuria and related
RL hyperphenylalaninemia: mutations and polymorphisms in the human
RL phenylalanine hydroxylase gene.";
RL Hum. Mutat. 1:13-22(1992).
RL [9]
RL DATABASE OF PKU VARIANTS.
RL MEDLINE=96174613; PubMed=8594560;
RL Hoang L., Byck S., Prevost L., Scriber C.R.;
RL "PAH Mutation Analysis Consortium Database: a database for disease-
RL producing and other allelic variation at the human PAH locus.";
RL Nucleic Acids Res. 24:127-131(1996).
RL [10]
RL VARIANT PRO-311.
RL MEDLINE=88294030; PubMed=2840952;
RL Lichter-Konecki U., Konecki D.S., Dilella A.G., Brayton K., Marvit J.,
RL Hahn T.M., Trefz F.K., Woo S.L.C.;
RL "Phenylalanine hydroxylase deficiency caused by a single base
RL substitution in an exon of the human phenylalanine hydroxylase
RL gene.";
RL Biochemistry 27:2881-2885(1988).
RL [11]
RL VARIANT LYS-280;
RL MEDLINE=89190664; PubMed=2564729;
RL Lyonnet S., Caillaud C., Rey F., Bertelson M., Frezal J., Rey J.,
RL Wunnich A.;
RL "Molecular genetics of phenylketonuria in Mediterranean countries: a
RL mutation associated with partial phenylalanine hydroxylase
```

RT RA deficiency.";

RT Am. J. Hum. Genet. 44:511-517 (1989).

RT [12]

RN VARIANT PRO-311.

RX MEDLINE=90136055; PubMed=2615649;

RA Hofman K.J., Antonarakis S.B., Missiou-Tsangarakis S., Boehm C.D.,

RA Valle D.;

RT "Phenylketonuria in the Greek population. Haplotype analysis of the

RT phenylalanine hydroxylase gene and identification of a PKU

RT mutation.";

RT Mol. Biol. Med. 6:245-250 (1989).

RN [13]

RN VARIANT LEU-364 DEL.

RX MEDLINE=90368081; PubMed=1975559;

RA Svensson E., Andersson B., Hagenfeldt L.;

RT "Two mutations within the coding sequence of the phenylalanine

RT hydroxylase gene.";

RL Hum. Genet. 85:300-304 (1990).

RN [14]

RN VARIANT GLN-261.

RX MEDLINE=91150775; PubMed=1671810;

RA Dianzani I., Forrest S.M., Camaschella C., Saglio G., Ponzone A.,

RA Cotton R.G.;

RT "Screening for mutations in the phenylalanine hydroxylase gene from

RT Italian patients with phenylketonuria by using the chemical cleavage

RT method: a new splice mutation.";

RL Am. J. Hum. Genet. 48:631-635 (1991).

RN [15]

RN VARIANT SER-255.

RX MEDLINE=91196738; PubMed=2014802;

RA Hofman K.J., Steel G., Kazazian H.H., Valle D.;

RA "Phenylketonuria in U.S. blacks: molecular analysis of the

RT phenylalanine hydroxylase gene.";

RL Am. J. Hum. Genet. 48:791-798 (1991).

RN [16]

RN VARIANTS TRP-252 AND LEU-281.

RX MEDLINE=91169506; PubMed=16722290;

RA Dworniczak B., Grudka K., Stumper J., Bartholome K.,

RA Aulehla-Scholz C., Horst J.;

RT "Phenylalanine hydroxylase gene: novel missense mutation in exon 7

RT causing severe phenylketonuria.";

RL Genomics 9:193-199 (1991).

RN [18]

RN VARIANTS SER-48 AND GLY-221.

RX MEDLINE=91348682; PubMed=1679030;

RA Konecki D.S., Schlöter M., Trefz F.K., Lichter-Konecki U.;

RT "The identification of two missense mutations at the PAH gene locus

RT in a Turkish patient with phenylketonuria.";

RL Hum. Genet. 87:389-393 (1991).

RN [19]

RN VARIANT PKU ILE-94 DEL.

RX MEDLINE=91236693; PubMed=1709636;

RA Callaud C., Lyonnet S., Rey F., Melle D., Frebourg T., Berthelon M.,

RA Vilarinho L., Vaz Osorio R., Rey J., Munnich A.;

RT "A 3-base pair in-frame deletion of the phenylalanine hydroxylase

RT gene results in a kinetic variant of phenylketonuria.";

RL J. Biol. Chem. 266:9351-9354 (1991).

RN [20]

RN VARIANTS NON-PKU HPA VAL-306 AND ASN-415.

RX MEDLINE=93052278; PubMed=1358789;

RA Economou-Petersen E., Henriksen K.F., Guldberg P., Guettler F.;

RT "Molecular basis for nonphenylketonuria hyperphenylalaninemia.";

RL Genomics 14:1-5 (1992).

RN [21]

RN VARIANTS PKU GLN-408 AND TRP-408.

RX MEDLINE=92380641; PubMed=1355066;

RA Lin C.H., Hsiao K.J., Tsai T.F., Chao H.K., Su T.S.;

RT "Identification of a missense phenylketonuria mutation at codon 408

RT in Chinese.";

RL Hum. Genet. 89:593-596 (1992).

RN [22]

RN VARIANT 364-LEU--GLU-368 DEL.

RX MEDLINE=93258322; PubMed=1363837;

RA Jaruzelska J., Melle D., Matuszak R., Borski K., Munnich A.;

RT "A new 15 bp deletion in exon 11 of the phenylalanine hydroxylase

RT gene in phenylketonuria.";

RL Hum. Mol. Genet. 1:763-764 (1992).

RN [23]

RN VARIANT LEU-244.

RX MEDLINE=93258323; PubMed=1363838;

RA Desviat L.R., Perez B., Ugarte M.;

RT "A new PKU mutation associated with haplotype 12.";

RL Hum. Mol. Genet. 1:765-766 (1992).

RN [24]

RN VARIANTS PKU.

RX MEDLINE=94010878; PubMed=8406445;

RA Guldberg P., Henriksen K.F., Guettler F.;

RT "Molecular analysis of phenylketonuria in Denmark: 99% of the

RT mutations detected by denaturing gradient gel electrophoresis.";

RL Genomics 17:141-146 (1993).

RN [25]

RN VARIANT GLY-190.

RX MEDLINE=93258345; PubMed=8098245;

RA Abadie V., Jaruzelska J., Lyonnet S., Millasseau P., Berthelon M.,

RA Rey F., Munnich A., Rey J.;

RT "Illegitimate transcription of the phenylalanine hydroxylase gene in

RT lymphocytes for identification of mutations in phenylketonuria.";

RL Hum. Mol. Genet. 2:31-34 (1993).

RN [26]

RN VARIANT VAL-276.

RX MEDLINE=94172999; PubMed=8068076;

RT Query Match 11.3%; Score 213.5; DB 1; Length 452;

RT Best Local Similarity 26.6%; Pred. No. 1.4e-09;

RT Matches 62; Conservative 44; Mismatches 100; Indels 27; Gaps 4;

QY 104 RNLVRLSSRFSLWKSICYCFRFFLDYLEAFGLLSDFLDH-----QAVIKFELETHF 155

DB 184 KKTGTVFKTLKSLYKTHA---CYEYNHIFPLEKYCGFHEDNIPQLEDVSOFLQCTG 240

QY 156 SYTPVSGFVAPHQVLSLLQDRVPPTASVNMRTLDKDNFSLTDLIHLHGHVPLHPSFS 215

DB 241 RLRPVAGLLSSRDFLGLAFRVFCHTQYIRHGSKPMYTPEDICHILLGHVPLFSRDSA 300

QY 216 EFFTMMGRLFTKVIKVOALPSKKORIOTLQSNLIAIVRCFWFTVSGLIENHEGRKAYG 275

DB 301 QFSQEIG-----LASLGAPDEVIEKLAi-----IYFTVFGCKQGSIKAYG 344

QY 276 AVLSSSQEILCHAFIDNVRLPLEIDDIIRLPNTSTPOETLFSIRHFDLVE 328

DB 345 AGLSSRFGELQYCLSERKPLLELEKTAIQNTVTPEPQPLYVAESFNDAKE 397

RESULT 21

PH4H MOUSE

ID PH4H MOUSE STANDARD; PRT; 453 AA.

AC P16331;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-AUG-1990 (Rel. 15, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Phenylalanine-4-hydroxylase (EC 1.14.16.1) (PAH) (Phe-4-

DN monooxygenase).

GN PAH.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OC NCBI\_TaxID=10090;

RN [1]

RN SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=LIVER;  
 RX MEDLINE=90241147; PubMed=2334400;  
 RA Ledley F.D.; Genet H.E.; Dundar B.S.; Woo S.L.C.;  
 RT "Mouse phenylalanine hydroxylase. Homology and divergence from human  
 phenylalanine hydroxylase.";  
 RL Biochem. J. 267:399-406(1990).  
 RN [2]  
 RP SEQUENCE OF 12-21.  
 RX MEDLINE=80220293; PubMed=7387651;  
 RA Weiborn M.; Humble E.; Ragnasson U.; Engstrom L.;  
 RT "Amino acid sequence at the phosphorylated site of rat liver  
 phenylalanine hydroxylase and phosphorylation of a corresponding  
 synthetic peptide.";  
 RL Biochem. Biophys. Res. Commun. 93:403-408(1980).  
 RN [3]  
 RP SEQUENCE OF 277-294.  
 RX MEDLINE=85122617; PubMed=6098294;  
 RA Robson K.J.H., Beattie W., James J., Cotton R.C.H., Morgan F.J.,  
 RA Woo S.L.C.;  
 RT "Sequence comparison of rat liver phenylalanine hydroxylase and its  
 cDNA clones.";  
 RL Biochemistry 23:5671-5675(1984).  
 RN [4]  
 RP VARIANTS PAH-ENU1 ALA-106 AND PAH-ENU2 SER-263.  
 RX MEDLINE=9224471; PubMed=9119379;  
 RA McDonald J.D.; Charlton C.K.;  
 RT "Characterization of mutations at the mouse phenylalanine hydroxylase  
 locus.";  
 RL Genomics 39:402-405(1997).  
 RL -1- CATALYTIC ACTIVITY: L-phenylalanine + tetrahydrobiopterin + O(2) =  
 L-tyrosine + dihydrobiopterin + H(2)O.  
 CC -1- COFACTOR: FERROUS ION.  
 CC -1- ENZYME REGULATION: N-TERMINAL REGION OF PAH IS THOUGHT TO CONTAIN  
 ALLOSTERIC BINDING SITES FOR PHENYLALANINE AND TO CONSTITUTE AN  
 "INHIBITORY" DOMAIN THAT REGULATES THE ACTIVITY OF A CATALYTIC  
 DOMAIN IN THE C-TERMINAL PORTION OF THE MOLECULE.  
 CC -1- PATHWAY: Catabolism of phenylalanine; first (rate-limiting) step.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- DISEASE: MOUSE STRAINS DEFICIENT IN PHENYLALANINE HYDROXYLASE  
 (PAH) WERE CREATED AS MODELS OF PHENYLKETONURIA (PKU).  
 CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID  
 HYDROXYLASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X51942; CAA36205.1; -  
 DR PIR; S15758; S15758.  
 DR HSSP; P00439; 2PAH.  
 DR MGD; MGI:97473; Pah.  
 DR InterPro; IPR002912; ACT.  
 DR InterPro; IPR001273; Aaa\_hydroxylase.  
 DR Pfam; PF00351; bioperin\_H; 1.  
 DR Pfam; PF01842; ACT; 1.  
 DR PRINTS; PR00372; FYMHDXLASE.  
 DR TIGRFAMs; TIGR01268; Phehydrox letr; 1.  
 DR PROSITE; PS00367; BIOTERIN\_HYDROXYL; 1.  
 DR OXIDOREDUCTASE; Monooxygenase; Phosphorylation;  
 KM Phenylalanine catabolism; Iron; Disease mutation;  
 FT MOD\_RES 16 16 PHOSPHORYLATION (BY PKA) (BY SIMILARITY).  
 FT METAL 285 285 IRON (BY SIMILARITY).  
 FT METAL 290 290 IRON (BY SIMILARITY).  
 FT METAL 330 330 IRON (BY SIMILARITY).  
 FT VARIANT 106 106 V -> A (IN PAH-ENU1; MILD PKU PHENOTYPE).  
 FT VARIANT 263 263 P -> S (IN PAH-ENU2; SEVERE PKU  
 PHENOTYPE).  
 FT METAL 154 154 IRON (POTENTIAL).  
 FT METAL 159 159 IRON (POTENTIAL).  
 SQ SEQUENCE 453 AA; 51928 MW; FAICDADDA598B8D9 CRC64;

Query Match 11.2%; Score 212.5; DB 1; Length 453;  
 Best Local Similarity 29.9%; Pred. No. 1.7e-09;  
 Matches 61; Conservative 35; Mismatches 87; Indels 21; Gaps 5;  
 QY 104 RNLM---YRLSSRSLWKSVCPRFFLDYLEAF-GLLSDFLDH-QAVIKFELETHFSYX 158  
 DB 184 RKTWGVFPRFLKALYKTHACVHNHNPFLPELKYCGFREDDNPQLBDSVQFLQTCGFRLR 243  
 QY 159 PVSGFAVPQYLSLLQDRYFPFASVWRTLDKDNFSLTPDILHDLGHVFWMLHPSPSEFF 218  
 DB 244 PVAGLLSDRDFLGLAFRFVHCTQYIRSGSKMPTPEPDICHELLGHVFLSDSPSPAGFS 303  
 QY 219 INMGRLFTVKIRKVOALPSKORIQTLQSNLAIIVRCFMTVESGLIENHKKYGAVAL 278  
 DB 304 QEIG-----LASLQAPDEVIEKLAT-----IYMTVFGLCKSGDSISKVAGGL 347  
 QY 279 ISSPOLGHAFIDNVRVPLFLDDQ 302  
 DB 348 LSSFGELQYCLSDKXKLPFLBLER 371  
 RESULT 22  
 ID PAH\_RALSO STANDARD; PRT; 313 AA.  
 AC Q8XU39;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Phenylalanine-4-hydroxylase (EC 1.14.16.1) (PAH) (Phe-4-  
 DE monooxygenase).  
 GN PHA OR RSC3355 OR RS02630.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 OC Ralstonia.  
 OX NCBI\_Taxid=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GM11000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salmoubat M.; Genin S.; Artiguenave F.; Gouzy J.; Mangenot S.;  
 RA Alier M.; Billaut A.; Broctier P.; Camus J.C.; Catolico L.;  
 RA Chandler M.; Choise N.; Claudel-Renard C.; Cunac S.; Demange N.;  
 RA Gaspier P.; Lavie M.; Moisan A.; Robert C.; Saurin W.; Schiek T.;  
 RA Schuster P.; Thebaud P.; Whalen M.; Wincker P.; Levy M.;  
 RA Weisenbach J.; Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";  
 RL Nature 415:497-502(2002).  
 CC -1- CATALYTIC ACTIVITY: L-phenylalanine + tetrahydrobiopterin + O(2) =  
 L-tyrosine + dihydrobiopterin + H(2)O.  
 CC -1- COFACTOR: Binds 1 ferrous ion (by similarity).  
 CC -1- PATHWAY: Catabolism of phenylalanine; first (rate-limiting) step.  
 CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID  
 HYDROXYLASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AL646074; CAD17143.1; -  
 DR InterPro; IPR001273; Aaa\_hydroxylase.  
 DR Pfam; PF00351; bioperin\_H; 1.  
 DR TIGRFAMs; TIGR01267; Phehydrox mono; 1.  
 DR PROSITE; PS00367; BIOTERIN\_HYDROXYL; 1.  
 KM Oxidoreductase; Monooxygenase; Phenylalanine catabolism; Iron;  
 FT Complete proteome 154 154 IRON (POTENTIAL).  
 FT METAL 154 154 IRON (POTENTIAL).  
 FT METAL 159 159 IRON (POTENTIAL).  
 SQ SEQUENCE 313 AA; 34901 MW; 2EBA9E216476371B CRC64;

```

Query Match          10.6%; Score 201; DB 1; Length 313;
Best Local Similarity 26.2%; Pred. No. 8.6e-09;
Matches 64; Conservative 41; Mismatches 95; Indels 44; Gaps 8;

QY 108 YRLSSRFSLWKSVCPR-----FFLDYLEAFGLSDFLDHOAVIKFFPEL-----152
DB 50 HRYTAADHATWRTLYDROEALLPGRACDFLOGLSTLG-----MSREGVSPFDRLNETILM 104
QY 153 -TFSTYPSVSGFVAPHQYLSLQDRYFFIASVMTLKDKNFSLTDLIHLGLHVPWILL 210
DB 105 RATGQIQIVAVGLVDPDEFVFEHLANRRFPASWMMRRPDQLDYLQEPDGFHDFGHVPLLI 164
QY 211 HPSISEFF--INMGRLFTKVIKQVQALPSKKORIOTLQSNLIAIVRCFWFTVESGLIENH 268
DB 165 NVPFADYMQAVGQGL-----KAARLGCALD-----MLARLYWTVVEFGLIRTP 207
QY 269 EGRKAYGAVLISSPQELGHAFIDNV--RVLPLELDQIIRLPFNSTPQETLFSIRHFDEL 326
DB 208 AGLRIYGAGIVSSKSESVA-LDSASPNRIGFDVHRINWTRYRIDTFQKTYFVIDSFQEL 266
QY 327 VELT 330
DB 267 FDAT 270

RESULT 23
PH4H_CHRVO STANDARD; PRT; 297 AA.
ID PH4H_CHRVO STANDARD; PRT; 297 AA.
AC P30967; Q9XC88; Q9R634;
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Phenylalanine-4-hydroxylase (EC 1.14.16.1) (PAH) (Phe-4-
monooxygenase).
GN PH4H.
OS Chromobacterium violaceum.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
OC Chromobacterium.
OX NCBI_TaxID=536;
RN [1]_TaxID=536;
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-20.
RC STRAIN=ATCC 12540;
RX MEDLINE=92011593; PubMed=1655752;
RA Onishi A., Iizota L.J., Benkovic S.J.;
RT "Cloning and expression of Chromobacterium violaceum phenylalanine
hydroxylase in Escherichia coli and comparison of amino acid sequence
with mammalian aromatic amino acid hydroxylases."
RT J. Biol. Chem. 266:18454-18459(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12540;
RA Volner A., Nersissian A.M., Abu-Omar M.M.;
RT "Expression, isolation, and metal-dependent catalysis of phenylalanine
hydroxylase from Chromobacterium violaceum."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: L-phenylalanine + tetrahydrobiopterin + O(2) =
L-tyrosine + dihydrobiopterin + H(2)O.
CC -!- COFACTOR: BINDS 1 COPPER ION.
CC -!- PATHWAY: Catabolism of phenylalanine; first (rate-limiting) step.
CC -!- SUBUNIT: MONOMER.
CC -!- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID
HYDROXYLASES FAMILY.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
frameshift in position 172.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; M55915; AAA23115.1; ALT_FRAME.
DR EMBL; AF146711; AAD37774.1; -.
DR PIR; A40996; A40996.
DR HSSP; P00439; 4PAH.
DR InterPro; IPR001273; Aaa hydroxylase.
DR Pfam; PF00351; biopterin_H; 1.
DR PRINTS; PR00372; FWHYDRXLASE.
DR TIGRFAMs; TIGR01267; Phe4hydrox mono; 1.
DR PROSITE; PS00367; BIOPTERIN HYDROXYL; 1.
KW Oxidoreductase; Monooxygenase; Phenylalanine catabolism; Copper.
FT METAL 138 COPPER (POTENTIAL).
FT METAL 143 COPPER (POTENTIAL).
SQ SEQUENCE 297 AA; 33594 MW; AAA67B6097171FB0 CRC64;

Query Match          10.4%; Score 196.5; DB 1; Length 297;
Best Local Similarity 27.4%; Pred. No. 1.8e-08;
Matches 63; Conservative 40; Mismatches 94; Indels 33; Gaps 6;

QY 126 FLDYLEAFGLSDFLDHOAVIKFFPELTFHSYFVSFVAPHQYLSLQDRYFPIASVM 184
DB 63 FLEGLELEVDADRVDPDFNKINEKLMATGKIVAVPGLIPDDVFFEHANRRFPVTWLL 122
QY 185 RTLDKNFSLTDLIHLGLHVPWLLHPSFSEFFINMGRLFTKVIKQVQALPSKKORIOT 244
DB 123 REPHQLDYQEPDVFHDLFGHVPLLINPVFADYLEAYKGGVKA-KALGALP-----173
QY 245 LQSNLIAIVRCFWFTVESGLIENHGRKAYGAVLISSPQ-----LGHAFIDNVRLPLEL 300
DB 174 -----MLARLYWTVVEFGLINTPAGMRTYAGAILSSKSESICYLDSASPNRV---GFDL 224
QY 301 DQIIRLPFNSTPQETLFSIRHFDELVELTSKLEWMLDQGLLESIPLYNQ 350
DB 225 NRIMNTRYRIDTFQKTYFVIDSFQELFDATP-----DFAPLYLQ 264

RESULT 24
PH4H_DROME
ID PH4H_DROME STANDARD; PRT; 452 AA.
AC P17276; Q27599; Q27600; O46110;
DT 01-AUG-1990 (Rel. 15, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein fenna [includes: Phenylalanine-4-hydroxylase (EC 1.14.16.1)
(PAH) (Phe-4-monooxygenase); Tryptophan 5-monooxygenase (EC 1.14.16.4)
(TRH) (Tryptophan 5-hydroxylase)].
GN HN OR TPH OR PAH OR CG7399.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S, and Oregon-R; TISSUE=Embryo, and Head;
RX MEDLINE=92156168; PubMed=1371286;
RA Neckameyer W.S., White K.;
RT "A single locus encodes both phenylalanine hydroxylase and tryptophan
hydroxylase activities in Drosophila."
RL J. Biol. Chem. 267:4199-4206(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91033030; PubMed=2121612;
RA Morales G., Requena J.M., Jimenez-Ruiz A., Lopez M.C., Ugarte M.,
Alonso C.;
RT "Sequence and expression of the Drosophila phenylalanine hydroxylase
mRNA."
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=96332435; PubMed=8769124;
RA Ruiz-Vazquez P., Moulard M., Silva F.J.;
RT "Structure of the phenylalanine hydroxylase gene in Drosophila

```





Db 347 LSSVGELEYCLTDKPKQKDFE-----PEVTGV---TKYPIOTQFQPLYYVADSFETAKE 396

QY 339 QGL--LESIP-----LYNQKYLSGFEVL 360

Db 397 KTIKANSIRPFQGVRYN--AYTQSVEVL 423

RESULT 25

PH4H CAUCR

ID PH4H CAUCR STANDARD; PRT; 294 AA.

AC Q9A7V7;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Phenylalanine-4-hydroxylase (EC 1.14.16.1) (PAH) (Phe-4-monooxygenase).

GN PH4H OR CC1612.

OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;

OC Caulobacter

OX NCBI\_TaxID=155892;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 19089 / CB15;

RX MEDLINE=21173699; PubMed=11259647;

RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E., Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R., Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B., DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H., Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K., Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O., Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;

RA "Complete genome sequence of Caulobacter crescentus.";

RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).

RL CC

CC -!- CATALYTIC ACTIVITY: L-phenylalanine + tetrahydrobiopterin + O(2) = L-tyrosine + dihydrobiopterin + H(2)O.

CC -!- COPACTOR: Binds 1 ferrous ion (BY similarity).

CC -!- PATHWAY: Catabolism of phenylalanine; first (rate-limiting) step.

CC -!- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID HYDROXYLASES FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; AB005836; AAK23591.1; -.

DR HSSP; P04176; 1PHZ.

DR TIGR; CC1612; -.

DR InterPro; IPR001273; Aaa hydroxylase.

DR Pfam; PF00351; bioprotein H; 1.

DR PRINTS; PR00372; FYWHYDRXLASE.

DR TIGRFAMs; TIGR01267; Phe4hydrox mono; 1.

DR PROSITE; PS00367; BIOPTERIN HYDROXYL; 1.

DR Oxidoreductase; Monooxygenase; Phenylalanine catabolism; Iron; Complete proteome.

KW

FT METAL 129 129 IRON (POTENTIAL).

FT METAL 134 134 IRON (POTENTIAL).

FT SEQUENCE 294 AA; 32761 MW; A632276FD0506720 CRC64;

QY 118 MKSYCPFFLDYLEAFGLSDFLDHQAVIKFFE-----LETHFSYYP----- 159

Db 24 WETVQAEHDVWITLYERQDTMLHGRACDEFWRGIDALDHRSGIDPFARINBELKRLTG 83

QY 160 -----VSGFVAPHQYLSLLQDRYFFPTASVMRTLDKDNFSLPDDLIHDLGHVPWLLHPSF 214

Db 84 WTVAVFGLVPDDVFFDHLANRRFPAGQFIRKPEHLDYLOQPDIFHDVFGHVPMLTDPVF 143

QY 215 SEFFINMGRFLTKVIEKQALPSKKORIQTLOSNIILAIIVRCFWFTVESGLIENHEGRKAY 274

Db 144 ADY-----MQAYGEGGRALGL-GRLANLARLYWYTVFEGFLMNTFAGLRIV 188

QY 275 GAVLISSPQELGHAFID-NVRVLPFLDQIIRLPENTSTPQETLFSIRHFDDELVELT--- 330

Db 189 GAGIVSRTSIFALDDPSNRIQFDLVRVMTLYRIDFQQVYFVIDSIQTLOEVTLRD 248

QY 331 -----SKLEWMLDQGLLESP 346

Db 249 FGAIYERLASVSDIGVAEIVP 269

RESULT 26

TY3H CAEL

ID TY3H CAEL STANDARD; PRT; 524 AA.

AC P09086;

DT 15-JUN-1998 (Rel. 36, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Tyrosine 3-monooxygenase (EC 1.14.16.2) (Tyrosine 3-hydroxylase) (TH).

GN CAT-2 OR B0432.5.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Henkhaus J., Wohldmann P.;

RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.

RL [2]

CC REVISIONS

RA Waterston R.;

RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.

CC -!- CATALYTIC ACTIVITY: L-tyrosine, tetrahydropteridine + O(2) = 3,4-dihydroxy-phenylalanine + dihydropteridine + H(2)O.

CC -!- COPACTOR: FERROUS ION.

CC -!- PATHWAY: Catecholamine biosynthesis; first step.

CC -!- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID HYDROXYLASES FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; U08036; AAB37888.2; -.

DR HSSP; P04177; 1TOH.

DR WormPep; B0432.5; CE29946.

DR InterPro; IPR001273; Aaa hydroxylase.

DR Pfam; PF00351; bioprotein H; 1.

DR PRINTS; PR00372; FYWHYDRXLASE.

DR PROSITE; PS00367; BIOPTERIN HYDROXYL; 1.

DR Catecholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron; Neurotransmitter biosynthesis.

KW

FT METAL 352 352 IRON (BY SIMILARITY).

FT METAL 357 357 IRON (BY SIMILARITY).

FT METAL 397 397 IRON (BY SIMILARITY).

FT SEQUENCE 524 AA; 59593 MW; 32B106950A629802 CRC64;

QY 19 KLRQSLSLFFQNSQRAYSTPYSYRIILQENKEQA-LARH-----KCISILE 69

Db 152 QLIHSALLTQNHVAL-----TKFSIFAKKLSDKNQSIWFRPRHISELQCSKCITKYE 206

Query Match 9.4%; Score 178.5; DB 1; Length 524;

Best Local Similarity 24.3%; Pred. No. 9.8e-07;

Matches 77; Conservative 41; Mismatches 114; Indels 85; Gaps 11;

```

OY 70 FFKLLFVHLISLKNQRECGSTDMVAVSTPFENNLYVRLLSRFSIMKSCYCPRLFDY 129
D 207 -----PTDPRHGHGDVAVYARRKFLND-----QALEFGFGDEIGY-----VDY 246
OY 130 LE-----AFGLSDGF-LDHOAVI-----KPELEETH 154
D 247 TEENHATWKAIVKEKGLDHLHSHTCAVYRONKILIOEKVLTADRIPOIRDNKFLQCKTG 306
OY 155 FSYVPSGVAPRHQVSLDQRYFPAISVMRLTDNDNSLPDLIHDLGHVPLHSP 214
D 307 FELRPGGLISARDFLASLAFVFTTTLRHKKSPHISPEPDLIHDLGHVPEFSDPL 366
OY 215 SEFFINNGRLFTVIEKVOALPSKKORIQTLQSNLIAIVRCGFWTVESGLIENHEGRKAY 274
D 367 AQMSQDIG-----LMSLGASDEHIKLSI-----VWFIVEFGKEDGKIKAI 410
OY 275 GAVLISSPOELGHAFID 291
D 411 GAGLLSAYGELMHACSD 427

RESULT 27
PH4H CAEEL STANDARD; PRT; 457 AA.
AC P90925;
DT 15-JUL-1998 (Rel. 36, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative phenylalanine-4-hydroxylase (EC 1.14.16.1) (PHE-4-monooxygenase).
GN K08F8.4
OS Caenorhabditis elegans.
OC Bursarota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Feloderinae; Caenorhabditis.
OX NBI_Ltaxid=6239;
RN [1]
RF SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Smye R.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: L-phenylalanine + tetrahydrobiopterin + O(2) =
CC L-tyrosine + dihydrobiopterin + H(2)O.
CC -1- COFACTOR: FERROUS ION (BY SIMILARITY).
CC -1- PATHWAY: Catabolism of phenylalanine; first (rate-limiting) step.
CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z66497; CAA91286.1; -
DR HSSP; P04176; 1PHZ.
DR Wormpep; K08F8.4; CE21050.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR001273; Aaa_hydroxylase.
DR Pfam; PF00351; bioplerin_H; 1.
DR Pfam; PF01842; ACT; 1.
DR PRINTS; PR00372; FYMHYDRXLASE.
DR TIGRFAMs; TIGR01268; Phehydrox_lectr; 1.
DR PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
KW Hypothetical protein; Oxidoreductase; Monooxygenase;
KM Phenylalanine catabolism; Iron.
FT METAL 285 285 IRON (BY SIMILARITY).
FT METAL 290 290 IRON (BY SIMILARITY).
FT METAL 330 330 IRON (BY SIMILARITY).
SEQUENCE 457 AA; 52129 MW; 68365836DFECGDAF CRC64;

```

```

Query Match 9.3%; Score 175.5; DB 1; Length 457;
Best Local Similarity 25.5%; Pred. No. 1,4e-06;
Matches 56; Conservative 35; Mismatches 80; Indels 49; Gaps 5;

OY 143 QAVIKFPELEETHFSYVPSGVAPRHQVSLDQRYFPAISVMRLTDNDNSLPDLIHDL 202
D 228 QDVSDPLKDCSTGYTIRPAGLLSSRDPLAGLAFVHFSHTQYIRHNSAPKYTEPDI CHEL 287
OY 203 LGHVPLHLPSPSEPFIMGRFLF-----TKYIEKVOALPSKKORIQTLQSNLIAIVRCFWF 258
D 288 LGHVPLFAVVERVQSGEIGLASLQADPDVIEKATL-----YWF 327
OY 259 TVESGLIENHEGRKAYGAVLISPOELGHAFIDNVRLPLE--LDQIRLPRNTSPQET 316
D 328 TIEFGICQDGRKKAIGAGLLSFGELQYALSDPEVWDPRVAVCCVTIKPTTEYQPKTF 387
OY 317 L-----FSIRH--FDELVELTSKL 333
D 388 LAESFASAKNKKLSMAATINRPFQIRYNAVYORVELIDKY 427

RESULT 28
NINQ DROME STANDARD; PRT; 1501 AA.
AC P10676; P10677;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neither inactivation nor afterpotential protein C (EC 2.7.1.37).
GN NINAC.
OS Drosophila melanogaster (fruit fly).
OC Insecta; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Preyptera; Neoptera; Endopterygota; Diptera; Brachycera;
OX Muscomorpha; Phlebotomidae; Drosophilidae; Drosophila.
OX NBI_Ltaxid=7227;
RN [1]
RF SEQUENCE FROM N.A.; FUNCTION, AND ALTERNATIVE SPLICING.
RX MEDLINE=88151067; PubMed=2449973;
RA Montell C.; Rubin G.M.;
RL "The Drosophila ninaC locus encodes two photoreceptor cell specific
RT proteins with domains homologous to protein kinases and the myosin
RT heavy chain head."
RL Cell 52:757-772(1988).
CC -1- FUNCTION: REQUIRED FOR PHOTORECEPTOR CELL FUNCTION. THE NINAC
CC PROTEINS JOIN PUTATIVE SERINE/THREONINE-PROTEIN KINASE AND MYOSIN
CC ACTIVITIES.
CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -1- SUBCELLULAR LOCATION: CYTOSKELETON.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; a long form (shown here) and a
CC short form; are produced by alternative splicing.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE SER/THR
CC FAMILY OF PROTEIN KINASES.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION, BELONGS TO THE MYOSIN
CC SUPERFAMILY.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03131; AAA28718.1; -
DR EMBL; J03131; AAA28719.1; -
DR EMBL; M20230; AAA28721.1; -
DR EMBL; M20231; AAA28720.1; -
DR PIR; A29813; A29813.
DR PIR; B29813; B29813.
DR HSSP; P08799; 1MND.
DR FlyBase; FBgn002938; ninaC.
DR InterPro; IPR000719; Euk_kinase.

```





```
Matches 72; Conservative 52; Mismatches 130; Indels 83; Gaps 14;
QY 30 NSQSQRAYSTPYVYRIILQENKEKQALAHKICISILEFFKRL--LFVHLLSLSKNQR 87
DB 773 NFNNILYDNDHLDNHNVLKCNVLEKQ-----CIQIAEFYKDLGSLVQTLQTLQDND- 825
QY 88 EGCSTDMNAVSTPFFNNRNWYRLLSRSFSLKSKYCPFFFDYLEAFGLSDFL-----DH 142
DB 826 ---STTVSLVET--FFNE-----FPKEFSF-----TLFEYLIKHKHLDLIFRFPQOH 868
QY 143 QAVTKFPELESTHESYYPVSGFVAPHQYLSLQDRYFPIASVMRTLDKNFSLTPOIHL 202
DB 869 DVLIQFFQESA-----PKYGHVANIQ--QILDGVSADAMNTLKNITVDDSKKGESL--- 917
QY 203 LGHPVWLLHPSFSEFFINNGRLFTKVIKQVQALPSKKQRIQTLQSLNLAIVFCWFTVES 262
DB 918 -----SECELHNAVKLSSLLVEK-----DNLDTNLRKIQVNLDTIDA 956
QY 263 GLIENHGRKAYGAVLISQBELGHAFIDNVRVLPLELDQIIRLPENTSTPQETLFSIRH 322
DB 957 E--KNISNKLKKEVQICRPNKGS-----IREVFNILVEELKSTTVN 998
QY 323 FDELVELTSKLEWMLDQGLLESIPLYNQEYKLSGFEV 359
DB 999 LSDVELVYSMLD--DEESLFIPLRLSLVDGNLLNPEV 1033

RESULT 32
RRPL EBOZM
ID RRPL EBOZM STANDARD; PRT; 2212 AA.
AC Q05318; Q39794;
DT 01-FEB-1994 (Rel. 28, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE RNA-directed RNA polymerase (EC 2.7.7.48) (Large structural protein)
DE (L protein).
GN L.
OS Ebola virus (strain Zaire Mayinga) (Ebo).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales; Filoviridae;
OC Ebola-like viruses.
OX NCBI_TaxID=128952;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99121703; PubMed=10073695;
RA Volchkov V.E., Volchkova V.A., Chepurinov A.A., Blinov V.M.,
RA Netesov S.V., Feldmann H.;
RT "Characterization of the L gene and 5' trailer region of Ebola
virus.";
RL J. Gen. Virol. 80:355-362(1999).
RN [2]
RP SEQUENCE OF 1-54 FROM N.A.
RX MEDLINE=94055391; PubMed=8237108;
RA Sanchez A., Kiley M.P., Holloway B.P., Auperin D.D.;
RT "Sequence analysis of the Ebola virus genome: organization, genetic
elements, and comparison with the genome of Marburg virus.";
RL Virus Res. 29:215-240(1993).
CC 1- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE ACTIVE POLYMERASE. IT
CC MAY FUNCTION IN RNA SYNTHESIS, CAPPING, AS WELL AS METHYLATION OF
CC CAPS, AND POLY(A) SYNTHESIS. ALSO INVOLVED IN TRANSCRIPTIONAL RNA
CC EDITING.
CC 2- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA) (N).
CC 3- SIMILARITY: PRIMARILY WITH THE N-TERMINAL HALF OF THE L PROTEINS
CC OF RHABDOVIRUSES AND PARAMYXOVIRUSES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR EMBL; X67110; CAA47483.1; -.
DR EMBL; AF068833; AAD14589.1; -.
DR EMBL; L11365; AAB81007.1; -.
DR InterPro; IPR001016; Viral_RNA_pol_L.
DR Pfam; PF00946; Paramyx_RNA_pol; 1.
KW Transferase; RNA-directed RNA polymerase.
SQ SEQUENCE 2212 AA; 252722 MW; 5B07EDDC370E2934 CRC64;

Query Match 5.3%; Score 100.5; DB 1; Length 2212;
Best Local Similarity 23.1%; Pred. No. 8.5;
Matches 62; Conservative 42; Mismatches 95; Indels 69; Gaps 13;

QY 13 ILKIALKRLQSLFQNSQSLQRAYSTPYVYRIILQENKEKQALAHKICISILEFFK 72
DB 355 LIRLEMPQQLCELF-----SIQKHGHPVLHSETAIQKVKHATVLKALRPVIFETV- 408
QY 73 NLLFVHLLSLSKNORESCGCTDMNVVS---TP-----FENRNLMYRLLSRSFSLKSKYCP 124
DB 409 ---CVFKYSIAKHVFSQCSWYSVTSDRNLTPGLNSYIKRNPPLPMIKELLWEFY--- 462
QY 135 FFIDYLEAFGL-----LSDFLDHQAIV-----KFELETHSFYSYPSGVAPHQYLSL 172
DB 463 -HLDPPLPFTKIIISLSIFIKDRATAVERTCWDVAFEPN-----VLGINPPHPSFK 514
QY 173 LQDRYFPIASVMRTLDKNFSLTPOIHLHGHVPWL--LHPFSFSEFF-----INMR 223
DB 515 RVPEQF-----LEQENFS-----IENVLSYAKLEYLLFQYRNFSLKEKELNVGR 561
QY 224 LFTKVIKQVQALPSKKQRIQTLQSLNLA 251
DB 562 TFGK-----LPYPTRNVTLCALLA 582

RESULT 33
Y075 MYCGE
ID Y075 MYCGE STANDARD; PRT; 1024 AA.
AC P47321; Q49190;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG075.
GN MG075.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bitt K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 1-156; 269-402; 643-736 AND 808-947 FROM N.A.
RX STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bitt K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing";
RL J. Bacteriol. 175:7918-7930(1993).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

CC EMBL; U39688; AAC71293.1; -  
 DR EMBL; U01715; AAC43189.1; ALT\_INIT.  
 DR EMBL; U02251; AAD12514.1; -  
 DR EMBL; U01749; AAD10562.1; -  
 DR EMBL; U01775; AAD10595.1; -  
 DR TIGR; MG075; -  
 KM Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 11 31 POTENTIAL.  
 FT TRANSSEM 951 971 POTENTIAL.  
 SQ SEQUENCE 1024 AA; 116424 MW; 5B15406855CB554 CRC64;

Query Match 5.2%; Score 98.5; DB 1; Length 1024;  
 Best Local Similarity 19.9%; Pred. No. 4.6;  
 Matches 68; Conservative 41; Mismatches 134; Indels 99; Gaps 11;

OY 29 QNSQSLQAVSTPYSYRIILQENKCK--QALARRKCSILEFFKRLVPHLLSLSKN 85  
 DB 564 QQTSLKRLFSV---IGDILSETVNRKITLHAAVKNNELSLIVETASTLKIKHL----- 613  
 OY 86 QREGCTDMAVSTPFRNRLMYRLSLRSFSLMKSYCDRFLDYLEAFGLSDPLDHOAV 145  
 DB 614 -----NVQYKVLVDKFEIKNSFIK-----ELINFPDITKDIPT 647  
 OY 146 IK--FELETH-----FSYVYSGFVAPRHOYSLLDQRYFPPLASVMTLDKDNF 192  
 DB 648 IKKYLFESENYKTLRKKEENEGPRGYMAKFTVPGTFSANTYSAL-----DKT 686  
 OY 193 SLTPDLIHDLGHVPLHPSSEPFIMNGRLFT-----KYLEKVALPSKKRIQT 244  
 DB 699 KSIIDLADNHL-FGKSLIESVNDSDFTKINSFTLKHYGDNLNLPNTHSLITKNGVQI 757  
 OY 245 LQSLALIVRCFWFTVESGLIENHGRKAYGAVLISFOELG----- 286  
 DB 758 VNVVFHIDALLTAELOAVFSNPK-----FVKSPELSKSLFEVWKTIFENSVNOI 810  
 OY 287 ---HAFIDNVRLPLELDOIRLPNNTSPOETLFSIRHFD 324  
 DB 811 LKKEYTKDNLKFFPKADGSSRLPEFLDSKPDORVIPPAYVD 852

## RESULT 34

CP1 PANAR STANDARD; PRT; 492 AA.

AC Q27712;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cytochrome P450 2L1 (EC 1.14.14.1) (CYP2L1).  
 GN CYP2L1.  
 OS Panulirus argus (spiny lobster).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Crustacea;  
 OC Malacostraca; Decapoda; Eucarida; Decapoda; Pleocyemata;  
 OC Palinura; Palinuroidea; Palinuridae; Panulirus.  
 OX NCBI\_TaxID=6737;

RN RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.

RC MEDLINE=96201120; PubMed=8619632;

RX James M.O., Boyle S.M., Ripido-Rosenthal H.G., Smith W.C.,

RA Greenberg R.W., Silverick K.T.; Triphido-Rosenthal H.G., Smith W.C.,

RT "CDNA and protein sequence of a major form of P450, CYP2L, in the

RL hepatopancreas of the spiny lobster, Panulirus argus."

CC Arch. Biochem. Biophys. 329:31-38(1996).

CC -FUNCTION: EFFICIENT IN CATALYZING THE MONOOXYGENATION OF

CC BENZHEXYLAMINE, AMINOPYRINE, BENZO (A) PYRENE, PROGESTERONE, AND

CC TESTOSTERONE.

CC -CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +

CC oxidized flavoprotein + H(2)O.

CC -SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By

CC similarity).

CC -SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U44826; AAB03106.1; -

DR HSSP; P00179; 1DT6.

DR Interpro; IPR001128; Cytochrome\_P450.

DR Pfam; PF00067; P450; 1.

DR PRINTS; PR00385; P450.

DR PROSITE; PS00086; CYTOCHROME\_P450; 1.

KM Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;

FT BINDING 436 436 HEME (BY SIMILARITY).

FT SEQUENCE 492 AA; 56767 MW; F27EE702D09D5EE CRC64;

Query Match 5.1%; Score 97; DB 1; Length 492;  
 Best Local Similarity 19.9%; Pred. No. 2.4;  
 Matches 74; Conservative 56; Mismatches 106; Indels 136; Gaps 18;

OY 56 KQALARRKCSILEFFKRLVPHLLSLSKNREGCTDMAVSTPFRNRLMYRLSLSNF 115  
 DB 79 KTLASKFESDPRDYTKLF-----GENDVAVV---FSNGV----- 113  
 OY 116 SLMKSYCDRFLDYLEAGL---LSDFLDHOAVIKFELETHFSY---YFVS----- 161  
 DB 114 -MMQTH-KRFTLRQLRDJGKSRLEAIIQEAACLVQELKRTDOPMLPSINLAVLN 171  
 OY 162 ---GFVAPHOYSLLDQ--RFPPLASVMTLDKDNFSLTPDLIHDLGHVPL--HPS 213  
 DB 172 VIKKLVAADHRY--SLDDEGGQFTQLLTITDNNQGFAL---NLFFYLPLMLITPD 223  
 OY 214 FSEFPINMGRLLFTKYLEKVALPSKKRIQTLSNLALIVRCFWFTVESGLIENHGRKA 273  
 DB 224 FVKNMGVAVLRDGVCE-----LKDYMKT-----FKKHQ----- 253  
 OY 274 YGAVLISFOELGHAFF-----IDNVRLPL----- 298  
 DB 254 -ATLDSNPKDLDAVYLIDLERKEDPLSTNNITETVAIVMDFGAGTETSTMIRWTL 312  
 OY 299 -----ELDOIIRLPNNTSPOETLFSIRHDELVELTSKLEWMLDQGLLESIPL----- 347  
 DB 313 YLMKYPEVOAKIQREIDAAVRGTLPSEHKDKLAYEATIHVV--HRIVSLVPLGVSHY 370  
 OY 348 YNOEKYLSGFV 359  
 DB 371 TNGDTLADGYRL 382

## RESULT 35

EX58 CHLPN STANDARD; PRT; 1050 AA.

AC Q92767; G93581; G932P2;

DT 30-MAY-2000 (Rel. 35, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Exodeoxyribonuclease V beta chain (EC 3.1.11.5).

GN RCB OR CP000738 OR CP0007.

OS Chlamydia pneumoniae (Chlamydia pneumoniae).

CC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

CC NCBI\_TaxID=83558;

CC [1]

CC SEQUENCE FROM N.A.

CC STRAIN=CML029;

CC MEDLINE=9920606; PubMed=10192388;

CC Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,

CC Olinger L., Grimwood J., Davis R.W., Stephens R.S.;

CC "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";

CC Nat. Genet. 21:385-389(1999).

CC [2]

RP SEQUENCE FROM N.A.  
RX STRAIN=AR39;  
RA MEDLINE=20150255; PubMed=10684935; Gill S.R., Heidelberg J.F.,  
Read T.D., Brunham R.C., Shen C., Gail S.R., Heidelberg J.F.,  
White O., Hickey E.K., Peterson J., Utterback T., Berry K.,  
Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
Cwin M., Nelson M., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
pneumoniae AR39.";  
RL Nucleic Acids Res. 28:1397-1406 (2000).  
[3]  
RP SEQUENCE FROM N.A.  
RX STRAIN=J138;  
RA MEDLINE=20330349; PubMed=10871362;  
Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
Shiba T., Iehi K., Hattori M., Kuhara S., Nakazawa T.;  
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
from Japan and CWL029 from USA.";  
RL Nucleic Acids Res. 28:2311-2314 (2000).  
CC -!- FUNCTION: REQUIRED FOR EFFICIENT DNA REPAIR; IT CATALYZES THE  
UNWINDING OF DOUBLE-STRANDED DNA AND THE CLEAVAGE OF SINGLE-  
STRANDED DNA AND IT STIMULATES LOCAL GENETIC RECOMBINATION.  
CC ALL OF THESE ACTIVITIES REQUIRE CONCOMITANT HYDROLYSIS OF ATP  
(BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage (in the presence of  
ATP) in either 5' to 3' or 3' to 5'-direction to yield 5'-  
phosphooligonucleotides.  
CC -!- SUBUNIT: CONSIST OF THREE SUBUNITS; RECB, RECC AND RECD  
(BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY. UVDR SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; AE001655; AAD18877.1; -;  
DR EMBL; AE002164; AAF37903.1; -;  
DR EMBL; AP002547; BAA98945.1; -;  
DR HSSP; P09980; IUA.  
DR TIGR; CP0007; -;  
DR InterPro: IPR000212; Uvdr-helicase.  
DR Pfam; PF00580; Uvdr-helicase; 1.  
KW Hydrolase; Nuclease; Exonuclease; Endonuclease; Helicase; ATP-binding;  
KW DNA repair; Complete proteome.  
FT NP\_BIND 21 28 ATP (POTENTIAL).  
FT CONFLICT 142 142 N -> K (IN REF. 2).  
FT CONFLICT 182 182 V -> I (IN REF. 1).  
SQ SEQUENCE 1050 AA; 121056 MW; 6E2CD03C2AEB83B7 CRC64;  
  
Query Match 5.1%; Score 97; DB 1; Length 1050;  
Best Local Similarity 22.1%; Pred. No. 6.2;  
Matches 77; Conservative 51; Mismatches 117; Indels 104; Gaps 19;  
  
QY 46 RIILQKENKQALAHKICISIL--EFFKNLLF---VHLLSLSKNQREGCGTDMA----- 95  
DB 137 RLHKNPALPHTSQVLVHHITNLYKLQDLKXNVLFGQFHLLAVRYNVTSKTSSLVVDKLLA 196  
QY 96 ----VSTPFNRR-----NLWV-RLSSRSLSKMSYCPFFFDLYLEA-----FG 134  
DB 197 SYTQPISSYFSSRVERLEQISLWHQIQVNSLLEIKP-----QVFLDQLTAHISGFKKQPS 252  
QY 135 LLSDFLDHQAIVKFELETH--FSYYPVS-GFVAPHQVLSLODRYFPAS----- 182  
DB 253 ILDDL--HHFVLLYTSETSSLSFSEFKIAETNFKRLA-----RYKPCATVLENWS 305  
QY 183 -VMTFLDKNF-----SLTDLTLHDL-LGHVPWLHLPSEFSSEFFINMRLTKVIEKQAL 235  
DB 306 WVERTLEPCNDRIFNTLLDLOEYLKQNTYTP-LSPDSVF-----ALEKLSS 354

QY 236 PSKORIQTQSLNLIIVRCFWFTVESGLIENHEGRKAYGAVLISSPQELGH--AFIDN 292  
DB 355 SEAPVQVAL-----REQYQLVLIDEFQDQTKQOWSIFSN 389  
QY 293 VRVPLELDQIIRLPENTSTPOETLFSIRHDELVELTSKLEWMLDOGL 341  
DB 390 LFISPKTKTGLS----FLIGDPKQSIYEWASADLPYLITAKSPFSEDKQL 434  
  
RESULT 36  
G6PD\_CHLPN STANDARD; PRT; 512 AA.  
AC Q928U6; Q9JQ22;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) (G6PD).  
GN ZWF OR CPN0238 OR CP0524 OR CPJ0238.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.   
OX NCBI\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CWL029;  
RX MEDLINE=99206606; PubMed=10192388;  
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
Olinger J., Grimwood J., Davis R.W., Stephens R.S.;  
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
RL Nat. Genet. 21:385-389 (1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AR39;  
RX MEDLINE=20150255; PubMed=10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
Cwin M., Nelson M., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
pneumoniae AR39.";  
RL Nucleic Acids Res. 28:1397-1406 (2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=J138;  
RX MEDLINE=20330349; PubMed=10871362;  
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
Shiba T., Iehi K., Hattori M., Kuhara S., Nakazawa T.;  
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
from Japan and CWL029 from USA.";  
RL Nucleic Acids Res. 28:2311-2314 (2000).  
CC -!- CATALYTIC ACTIVITY: D-glucose 6-phosphate + NADP(+) = D-glucono-  
1,5-lactone 6-phosphate + NADPH.  
CC -!- PATHWAY: Pentose phosphate pathway; first step.  
CC -!- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE  
FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; AE001609; AAD18391.1; -;  
DR EMBL; AE002211; AAF73682.1; -;  
DR EMBL; AP002545; BAA98448.1; -;  
DR HSSP; P11411; 1DPG.  
DR TIGR; CP0524; -;  
DR InterPro: IPR001282; G6PD.  
DR Pfam; PF00479; G6PD; 1.  
DR Pfam; PF02781; G6PD\_C; 1.

```

DR PRINTS: PRO0079; G6PDHRCGNASE.
DR FREDOM; PD001129; G6PD; 1.
DR TIGRNAME; TIGR00871; zwf; 1.
DR PROSITE; PS00069; G6P_DEHYDROGENASE; 1.
DR OXIDOREDUCTASE; NADP; Glucose metabolism; Complete proteome.
FT ACT_SITE 205 205 BY SIMILARITY.
SQ SEQUENCE 512 AA; 58739 MW; EABC8DBFC940BCE CRC64;

Query Match 5.1%; Score 96.5; DB 1; Length 512;
Best Local Similarity 18.8%; Pred. No. 2.7;
Matches 83; Conservative 57; Mismatches 124; Indels 177; Gaps 22;

QY 20 LROSLSLFFGNSQSLORAY-----STPYSTYRIILQENKEKQALARKICISILE 69
DB 113 LKDSL-----EDLKYTGTRGNRLFYLTPTPGYFSRIINLNKHK----- 152
QY 70 FFKNLFPVHLISLKNQSGCSTDAVAVSTP-----FNRLMY----- 108
DB 153 -----LFY-----KNQDGKPMRWIIIEKPGRLDSAKQLQCCINENINSYHIDH 201
QY 109 -----RLSSRF-----SLKSYCFRFLDYLF-----APGLSDPLDQAVIK 147
DB 202 YLQKRYQNLITFRFANTIFESCMNSQ-----YIDHVQSLSEITGIGSRGNFEEKGMRL 257
QY 148 -----FPELETHFSYPPVSGVAPR-----QYISLQDRYFPI-----ASVMR----- 185
DB 258 DWQONHMMQLCLLTWEPPTPDADIRKIKIKILO-NISPSGSSIVKQYGPVTGQ 316
QY 186 -----TLDKNFSLTPLDILHDLGHVFWLHPSFSEFFINMGRLEFTYIEKVOAL 235
DB 317 VSVLYGVEENVDKDSRVTYVALKTVINNPMGLVP-----FYRAKRLAKSTDISII 372
QY 236 PSKQRIQTLQSNLIAIVCFWFTVESGLIENHEGRKAYGVALISSPOLGHAFTDNVR- 294
DB 373 FKKS-----PYNLFAEBCSCRCPIENDL-----IIRIQPDE-GVALKNCKY 414
QY 295 -----VLPLEDOIIRLFPNTSTPO-----ETLFSIRHPELV 327
DB 415 PGTNNIVRPVKMDFRYDSYFOGTTTPEAYRRLDCIIGDRTLFTGDEVNAWMLFPPVL 474
QY 328 ELTSKLEMMLDGLESIPLY 348
DB 475 E-----EM--DQSSPSFPNY 488

RESULT 37
YABD_SCHPO STANDARD; PRT; 529 AA.
ID YABD_SCHPO
AC 009812;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C2G11.13 in chromosome 1.
GN SPAC2G11.13.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
ON NCBI_TaxId=4896;
RX NCBI_TaxId=4896;
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RX Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Delavault T., Frazer A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagers K.,
RA James K., Jones L., Jones W., Leather S., McDaniel S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,

```

---

```

RA Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tiver A., Walsh S.V., Warren T., Whitehead S.,
RA Woodard J., Volckaert G., Aert R., Robben J., Grynopret B.,
RA Welfjens I., Vanstreels E., Kieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Filicz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehner H., Reinhardt R., Pohl T.M.,
RA Rger P., Zimmermann W., Medler H., Wandut R., Fumelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Huret S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Rhode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Useery D., Barrell B.G., Nurse P.;
RA "The genome sequence of Schizosaccharomyces pombe." ;
RT Nature 415:871-880(2002).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potentia1).
CC -1- SIMILARITY: SOME. TO YEAST YCL38C AND B. SUBTILIS YX10.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z54354; CAA91178.1; -
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 25 45 POTENTIAL.
FT TRANSMEM 72 92 POTENTIAL.
FT TRANSMEM 130 150 POTENTIAL.
FT TRANSMEM 241 261 POTENTIAL.
FT TRANSMEM 269 289 POTENTIAL.
FT TRANSMEM 365 385 POTENTIAL.
FT TRANSMEM 397 417 POTENTIAL.
FT TRANSMEM 429 449 POTENTIAL.
FT TRANSMEM 467 487 POTENTIAL.
FT TRANSMEM 496 516 POTENTIAL.
SQ SEQUENCE 529 AA; 58641 MW; 18E2B0FAF8BD8F CRC64;

Query Match 5.1%; Score 96; DB 1; Length 529;
Best Local Similarity 22.0%; Pred. No. 3.1;
Matches 66; Conservative 35; Mismatches 127; Indels 72; Gaps 11;

QY 120 SYCPREFLYLAEAGLSD--FLDHQAVIKFPELETHFSYPPVSGVAAHQVLSLQDRY 177
DB 133 SMSPSFLLGKIVFLVDLNFITISQCYDSF-LPIFLRFPYLR--GITTESALQDET 189
QY 178 FPIASVMTLKDQNFSLTPDLI-HDLIGH-----VWLLHPSFSEFF-- 218
DB 190 DDDLSYITWTIDSEEPYLLSHSLINSEAPADVEDEHKAKIARLSVSGFSFGCA 249
QY 219 -INMGRLEFTKVIKQVALPSKQRIQTLQSNLIAIVCFWFTVESGL-----IENHEG 270
DB 250 AILFOIIFPIPIYKTNMNP-----IILPTVAVSCMWLILSTLCTIVTLVPEVNHSS 302
QY 271 RKAYGAVLSSPOLGHAFTDNVRVLPLEDOIIRLFPN-----TSRPOELFS----- 319
DB 303 -DAIFLLVNVKESYHSFKKAMSISIRLPLFRPLINCIGIQTSSSAVIFGKARLMS 361
QY 320 -----IRHPELVETLSKLEMMLDGLESIPLYNDEKLYSGE 358
DB 362 NPQUTLLGNGISSFALLGTVIIPVITEYFQNLNSQVVMISILLPMAPLYGLGIIPGR 421

RESULT 38
TBSU_HSV7J STANDARD; PRT; 2059 AA.
ID TBSU_HSV7J
AC P52362;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```





```

QY 224 LFT-----KYLEKQALPSKQ-----RIQTLQSLTAIVRCFWTVESGLIE--NHG 270
DB 799 TVGSGNNINFLPCGCGSGKRGKQASAAATITFTLQASTKTSINEYIDPILKYNBEG 858
QY 271 RK-----AYGAVLISSPQELIGHAF-----IDNVR-----VLFLEL 300
DB 859 QKIEPQYIPVITLIVNGCEGIGTGSFIPNYKQIIDIKYINKEPLIMVWYK 918
QY 301 DOIIRLPNTSTPQETLFSIRHPD-----ELVELTSKLEWMD--QGLLESIPLYNQEKY 353
DB 919 DFKRIESNGKTGYETIGIINKIDNDLTLETLPK-KWTQDYKEFLER--LTLDEKH 973

RESULT 40
SRB8 YEAST STANDARD; PRT; 1427 AA.
ID SRB8 YEAST STANDARD; PRT; 1427 AA.
AC P25648; P25647;
DT 01-MAY-1992 (Rel. 22, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Suppressor of RNA polymerase B SRB8.
CN SRB8 OR YCR081W OR YCR81W/YCR80W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95293223; PubMed=7774808;
RA Hengartner C.J., Thompson C.M., Zhang J., Chao D.M., Liao S.M.,
RA Koleske A.J., Okamura S., Young R.A.;
RT "Association of an activator with an RNA polymerase II holoenzyme.";
RL Genes Dev. 9:897-910(1995).
RN [2]
RP SEQUENCE OF 1-530 FROM N.A.
RA Bailestra J.P.G., Franco L., Hoenicka J., Jimenez A., Remacha M.,
RA Sanz E.;
RN Submitted (MAR-1992) to the EMBL/Genbank/DBJ databases.
RL [3]
RP SEQUENCE OF 531-1427 FROM N.A.
RA Feldmann H., Mannhaupt G., Vetter I.;
RN Submitted (MAR-1992) to the EMBL/Genbank/DBJ databases.
RN [4]
RP REVISIONS.
RA Gramada R.;
RL Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF THE RNA POLYMERASE II HOLOENZYME AND THE
CC MEDIATOR OF ACTIVATION SUBCOMPLEX.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X59720; CAA42268.1; -
DR PIR; S19496; S19496.
DR PIR; S19495; S19495.
DR TRANSFAC; T02152; -
DR SGD; S0000677; SRB8.
DR InterPro; IPR000651; RaseGFPN.
RW Nuclear protein.
SQ SEQUENCE 1427 AA; 166859 MW; 776A2A9D825331FE CRC64;

Query Match 5.0%; Score 95; DB 1; Length 1427;
Best Local Similarity 18.8%; Pred. No. 13;
Matches 98; Conservative 65; Mismatches 155; Indels 204; Gaps 25;

12 YILKALKROSLSLFQNSQSLQRAVSTPYSYRRI-----LQENKEKQALARRKCI 65
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

---

```

DB 214 YILE---KLIFPMTHNYSQOL--RTWKQISYFLKLGNCYSRLINKE---IFHHVLV 266
QY 66 SLEFPKNLFP-----VHLISLTKQREGCSIDMAVYST-----PFF----- 102
DB 267 EFLINKMENEFLPLSLHLMIFMDICQIDINAPVAATITTSQKEPFLVTKITMLLHK 326
QY 103 -----NENLWYRLSSRFP-----WKS 120
DB 327 YVIVSSKAMINDENYIINDIKKNNKIKLNLKLSLLIKIFQEGSLEVFIFPTSNMEI 386
QY 121 YCFRFF-----LDYLEAFGLSDPLDHOAVIKFPELTHESYVSGFVAP 166
DB 387 YKPLFELIVSNADTNQNSDMKKLELSYRNESIKNNSSIR-----NVIAMSANAN 437
QY 167 HOYLSLLODRYFPFIASV-----RTLQDN-----FSLTPDLIHDLGHVP 207
DB 438 DFLQTLVTCKOPFKLSQINCIDTQFTKLD--DNPTFPMPTVYQNDLTHKIIQLIL 496
QY 208 WLHPSS--PSEFFINMGRLEFKVLEKQVALPSKKORIQLQSLN--IAVRCFWFTVSG 263
DB 497 WSHPSRQPDHVESN-----QLVAKLLL-----RINSTDDEHBFQEDAINSLVWQ 544
QY 264 LIENHEGR-----AYGAV-----LISSP-----QELGHPID-- 291
DB 545 LAKPFAQKRVVSYNMPSLRYLNLITIGIKPYIRKLISSGLLYIQDSNDFVHQ 604
QY 292 ---NRAVLPLELDQIIRLPNTSTPQETLFSIRHDELVELTSK----- 333
DB 605 LLNLKISPLMKSYQNMVLRNVMEDVKEIFEFDVLVETIQIMRILSDITNLQLS 664
QY 334 -----EMLMD--QGLLESIPLYNQEKYSGFEVLQ 361
DB 665 KTLPLSIKIMVASEWYLSHLCGLSSV--NRTVLKLIPIKIFC 703

RESULT 41
TRA2 CAEEL STANDARD; PRT; 1475 AA.
ID TRA2 CAEEL STANDARD; PRT; 1475 AA.
AC P34709;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sex-determining transformer protein 2 precursor.
GN TRA-2 OR C15F1.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Gelodertinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Brictol N2;
RX MEDLINE=92360913; PubMed=1498366;
RA Kuwabara P.B., Okkema P.G., Kimble J.;
RT "tra-2 encodes a membrane protein and may mediate cell communication
RT in the Caenorhabditis elegans sex determination pathway.";
RL Mol. Biol. Cell 3:461-473(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Brictol N2;
RA Bentley D., Favetto A.;
RN Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP REVISIONS.
RA Waterston R.;
RL Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: PROMOTES FEMALE DEVELOPMENT IN XX ANIMALS WHERE IT
CC SEQUESTERS ONE OR MORE OF THE FEM PROTEINS TO THE MEMBRANE THEREBY
CC FREEING THE TRA-1 PROTEIN (A PUTATIVE TRANSCRIPTION FACTOR) TO
CC ENTER THE NUCLEUS AND PROMOTE FEMALE DEVELOPMENT. IN XO ANIMALS IT
CC ACTS AS A RECEPTOR FOR HER-1 WHICH PREVENTS IT FROM BINDING TO FEM
CC PROTEINS THEREBY REPRESSING THE ACTIVITY OF TRA-1.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: SOMATIC AND GERM LINE TISSUES.

```



Db 132 EANAATFPVDDAPVKLKPVTFLFKVDNDVFPKLNIPKGINISDNIAPFERNYKLSIEL 191  
QY 190 DNEFSLT-----PDLIHLDLGHVPLWLPSPSEPFIMKGLFTKVIKVOALPSSKKQRIQTL 245  
Db 192 KNFDLAFURKADVIITIKKQKRYNFBEMQOTYVGGGNTVINLDTLQAQVNFANIQDL 251  
QY 246 QSNLIAI-----VRCEFWFVESGLIEN-----HEGRKAYGAVLISSEFQELGHAFFIDNR 294  
Db 252 QNTFVKVGNLSTQLEFIPVKNLTDNAGNDLTHIAKTVIGSEFPQTNVNLAKSVIEYDK 311  
QY 295 VLPL---ELDQIIRLPP 308  
Db 312 VOPLVKQAFERRVLTFF 328  
RESULT 43  
ID Y065\_MYCGE STANDARD; PRT; 466 AA.  
AC P47311: 049281:  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical ABC transporter ATP-binding protein MG065.  
GN MG065.  
OS Mycoplasma genitalium.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OK NCBI\_TaxID=2097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33530 / G-37;  
RX MEDLINE=96026346; PubMed=7569933;  
RA Frieser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,  
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,  
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,  
RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,  
RA Tomb J.-F., Dougherty B.A., Bort K.F., Hu P.-C., Lucier T.S.,  
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;  
RT "The minimal gene complement of Mycoplasma genitalium";  
RL Science 270:397-403 (1995).  
[2]  
RN SEQUENCE OF 393-466 FROM N.A.  
RP STRAIN=ATCC 33530 / G-37;  
RX MEDLINE=94075220; PubMed=8253680;  
RA Peterson S.N., Hu P.-C., Bort K.F., Hutchison C.A. III;  
RT "A survey of the Mycoplasma genitalium genome by using random  
sequencing";  
RL J. Bacteriol. 175:7918-7930 (1993).  
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U39686; AAC71283.1; -  
DR EMBL: U02154; AAD12436.1; -  
DR TIGR: MG065; -  
DR InterPro: IPR003593; AAA\_ATPase.  
DR InterPro: IPR003439; ABC\_transporter.  
DR Pfam: PF00005; ABC\_tran.1.  
DR ProDom: PD000006; ABC\_transporter; 1.  
DR SMART: SMO0382; AAA; 1.  
DR PROSITE: PS00211; ABC\_TRANSPORTER; 1.  
KM Hypothetical protein; ATP-binding; Transport; Complete proteome.  
FT NP\_BIND 269 276 ATP\_BINDING [POTENTIAL].  
FT COMPLECT 394 399  
SQ SEQUENCE 466 AA; 54009 MW; A585044B8A90391C CRC64;  
Query Match 4.9%; Score 92.5; DB 1; Length 466;  
Best Local Similarity 22.0%; Pred. No. 5;

Matches 61; Conservative 44; Mismatches 95; Indels 77; Gaps 15;  
QY 68 LEFPKRLVPHLLSLSTKQREGCSTDAVAVSTPFERNIMYRLSSRPSLMK----- 119  
Db 31 LEESQNYLQKGLKQKKEYKSI-----IYNPLEK---RFKMKENTETFE 76  
QY 120 SYCPFFPLDYL-AFGILSDPLDQAVIKFELETHFSTYPSGVAPAHQYLSL----- 172  
Db 77 SYDRFFFTKYNHYSLSLPSFINEQ-----LEVIASV--NSPLNEHNKLAENKQSF 127  
QY 173 -LQDRYFPPIASVWRTLDKNFSLTPTD-----IHDLLGHVPLWLPSP 213  
Db 128 SFPEKLFQATQGFNNLEK-NTAIDSDLPLOFKVNTQLOKQREERKLNLIK-LKULS 185  
QY 214 -----FSEPFIMKGLFTK-VIEKVOALPSSKKQRIQTLQ--SNLIAIVRCFWFVES 262  
Db 186 EKKQETILNNWFSNERRFLKNEVKVWNLSPKQQAQIDQNIILKNVYKY-ITN 244  
QY 263 GLIENHEGR-----KAYGAVLISSEFQELGHAFFIDNR 293  
Db 245 GITTNVAKGVDLAKSHDFVILGPGSGGKTLTINI 281  
RESULT 44  
ID FLP\_KLULA STANDARD; PRT; 447 AA.  
AC F13783;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Recombinase FLP protein.  
GN A.  
OS Kluyveromyces fragilis (Yeast).  
OG Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetaceae; Saccharomycetaceae; Kluyveromyces.  
OK NCBI\_TaxID=28985;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86232585; PubMed=3520486;  
RA Chen X.J., Saltoia M., Falcone C., Bianchi M.M., Fukuhara H.;  
RT "Sequence organization of the circular plasmid pK01 from the yeast  
Kluyveromyces fragilis";  
RL Nucleic Acids Res. 14:4471-4481 (1986).  
CC -!- FUNCTION: FLP CATALYZES RECOMBINATION BETWEEN THE LARGE INVERTED  
CC REPEATS OF THE PLASMID.  
CC -----  
CC -!- SIMILARITY: BELONGS TO THE "PHAGE" INTEGRASE FAMILY.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X03961; CAA27591.1; -  
DR PIR: S28086; S28086.  
KM DNA recombination; DNA integration; Plasmid.  
FT ACT\_SITE 336 336 TRANSIENT COVALENT LINKAGE TO DNA DURING  
FT STRAND CLEAVAGE AND REJOINING (BY  
FT SIMILARITY).  
SQ SEQUENCE 447 AA; 51110 MW; 89D08AB52718A104 CRC64;  
Query Match 4.9%; Score 92; DB 1; Length 447;  
Best Local Similarity 23.6%; Pred. No. 5; 2; Mismatches 82; Indels 56; Gaps 12;  
Matches 54; Conservative 37;  
QY 13 ILKIALKROSLSPFQNSQSLQRAYSTPSTYRIILQK-ENKEKALAHKICSLTEFF 71  
Db 81 IIFKTVLRK-----TESLQKIDIESALPSTYKVVSPKNOEVLDFRYE-----ETH 127  
QY 72 K-----NLLFVHLTSLSKN-----QREGCSTDAVAVSTPFERNIMYRL-----SS 113

Db 128 KYDASWGLQNTILSKEKIDKIVIRIACFFDQSCVTT---TKRAEYRLLLLGAVGNCC 184  
 QY 114 RSLWKSVCPRFFLDYLEAGLLSDFLDHOAVIKFELETHFSYYPVSG----FVAPHQY 169  
 Db 185 RYSDKUNLDPRFTFIYNNP--LGPVIRATVTETKSRTERVNFVPGVNGDCDLLSLDYD 242  
 QY 170 LSLLDQRYEPIASVMTLKDKNFSLTP-DLIHDLGLGHVFWLLHPSFSEF 217  
 Db 243 L-----RVCSPIEKTVSSNRPNTQTHQFL---PESLARTFSRF 277

## RESULT 45

MATK ARATH  
 ID MATK ARATH STANDARD; PRT; 526 AA.  
 AC P56784;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Probable intron maturase (Maturase K).  
 GN MATK OR YCF14.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Chloroplast  
 OC Eukaryota, Viridiplantae, Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=cv. Columbia;  
 RC MEDLINE=20039611; PubMed=10574454;  
 RA Sato S., Nakamura Y., Kaneko T., Asamizu E., Tabata S.;  
 RT "Complete structure of the chloroplast genome of Arabidopsis  
 thaliana.";  
 RL DNA Res. 6:283-290(1999).  
 CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II  
 CC INTRONS.  
 CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,  
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED  
 CC BY MITOCHONDRIAL INTRONS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AP000423; BA084366.1; -.  
 DR InterPro; IPR000442; Intron\_mature2.  
 DR InterPro; IPR002866; MatK N.  
 DR Pfam; PF01348; Intron\_maturas2; 1.  
 DR Pfam; PF01824; MatK N; 1.  
 KW Chloroplast; mRNA processing.  
 SQ SEQUENCE 526 AA; 63038 MW; 4798486C56ACC011 CRC64;

Query Match 4.9%; Score 92; DB 1; Length 526;  
 Best Local Similarity 22.1%; Pred. No. 6.4;  
 Matches 65; Conservative 35; Mismatches 118; Indels 76; Gaps 11;

QY 2 HVCERTDPKYLKIALKRLQSLSLFFONSQSLQRAYSTPYVYRIILQKENKEKQALR 61  
 Db 236 HVCE-----YESIFFLKRSHLSTSYEVLFEIIVPYGK----- 271  
 QY 62 HKCISILEFFKNLL--FVHLLSLSKNQ-----RECGSTDMAVSTP-----FFRNILM 107  
 Db 272 -----IHHFFKVFVNNFPAILGLLKDPFHIVYRIGRCILATKPTLLMKWKYIFVNLW 326  
 QY 108 YLLSSRLSKWSCPRFFLDYLEAGLLSDFLDHOAVIKFELETHFSYYPVSGFVAPH 167  
 Db 327 QCYFSVWFQSQKVNINQLSKDNLEFJGYLLSLRLNPLVVRSMLE-----NSFLIDN 378

QY 168 QYLSLLQDRYPFIASVMTLKDKNFSLTDLIHLGLH-----VPMLLHPSFSEFFINMR 223  
 Db 379 VRIKL--DSKIPISIIIGSLAKDKFC-----NVLGHPIKATW-TDSSSDILNRFVR 428  
 QY 224 LFTKVIKVOALPKQKQ-----IOTLQSNLIAIVRCFWFTVVGSLIE 266  
 Db 429 ICRNISHYSGSGKKKNLYIRIKYLRLCCVKTARKHKSTVRTFLKRLGSLGLE 482

## RESULT 46

MATK MAIZE  
 ID MATK MAIZE STANDARD; PRT; 544 AA.  
 AC P48190;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Probable intron maturase (Maturase K).  
 GN MATK OR YCF14.  
 OS Zea mays (Maize).  
 OC Chloroplast  
 OC Eukaryota, Viridiplantae, Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC Clade;  
 OC Panicoideae; Andropogoneae; Zea.  
 OX NCBI\_TaxID=4577;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95395841; PubMed=7666415;  
 RA Maier R.M., Necker mann K., Igloi G.L., Koessel H.;  
 RT "Complete sequence of the maize chloroplast genome: gene content,  
 RT hotspots of divergence and fine tuning of genetic information by  
 RT transcript editing.";  
 RL J. Mol. Biol. 251:614-628(1995).  
 CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II  
 CC INTRONS.  
 CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,  
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED  
 CC BY MITOCHONDRIAL INTRONS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X86563; CAA60266.1; -.  
 DR MaizeDB; 118213; -.  
 DR InterPro; IPR000442; Intron\_mature2.  
 DR InterPro; IPR002866; MatK N.  
 DR Pfam; PF01348; Intron\_maturas2; 1.  
 DR Pfam; PF01824; MatK N; 1.  
 KW Chloroplast; mRNA processing.  
 SQ SEQUENCE 544 AA; 64800 MW; 95D6C38D39EC84D0 CRC64;

Query Match 4.9%; Score 92; DB 1; Length 544;  
 Best Local Similarity 18.6%; Pred. No. 6.7;  
 Matches 73; Conservative 64; Mismatches 140; Indels 116; Gaps 16;

QY 1 VHYCERTDPKYLKIALKRLQSLSLFFONSQSLQRAYSTPYVYRIILQKENKEKQALA 60  
 Db 199 LEYRQDVPSSLHLRLFFLYYSNNWSFITSMSKI-----FLKKENKELPRFL 246  
 QY 61 RHKCSILEFFKNLLFVHLLS-----LSKNQREGCSTDMAVSTPFFNRLWY- 108  
 Db 247 YNSYVSEYEFF--LLFLHKQSSCLRLTSSGTFLERIIIFSGKMEHFGVWVPGFFKRLIWF 304  
 QY 109 -----RLSSRESL-----WKSYPFRFLDYLEAFGLSDFLDHOAVIKFPE 150  
 Db 305 MDPLMHVYRQKAILASKGTLKLLKKWKSILVNFSSQVFFSWTQPORIRLNQLTNSCFD 364  
 QY 151 LETHFSYYPVSGFVAPHQYLSLQDRYFIASVMTLKDKNFSLTDLIHLGLGHVFWLL 210

DB 365 FLGYLSVPIINTLVLRNOML-----ENSLFDITRMKKFD-----ITVLATPLVGSGL----- 410

QY 211 HSESEFEFINMGRLETK-----VIEKVAL-----PSKKRIQTLSGLI 250

DB 411 --SKAOFCTGSGHPISKFWTDLSMDILDRPRICRNIFHHSSSKKQLYLTK----- 464

QY 251 AIVRCFWFTVESGLIENHEGRKAYGAVLISSPEGLHATIDNVRVLELDQIIRLPNT 310

DB 465 YLLR--LSCARTLARKK-----STVRTFWRLSGLVLEEP----- 498

QY 311 STPOETLFSIRHFDLVELT-----SKLEWMLD 338

DB 499 FTBEBOVPSLM-FTKTIHFPHSGHSECIMYLD 530

RESULT 47

APR HUMAN STANDARD; PRT; 4563 AA.

AC P04114; 000502; Q13787;

DT 01-NOV-1986 (Rel. 03, Created)

DT 01-NOV-1986 (Rel. 03, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Apolipoprotein B-100 precursor (Apo B-100) [Contains: Apolipoprotein B-48 (Apo B-48)].

GN AP08. Homo sapiens (Human).

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN (1)

RP SEQUENCE FROM N.A.

RX MEDLINE=8701385; PubMed=3763409; Pease R.J., Lusis A.J., Knott T.C., Wallis S.C., Powell L.M., Pease R.J., Lusis A.J., Blackhart B., McCarthy B.J., Mahley R.W., Levy-Wilson B., Scott J.; "Complete cDNA and derived protein sequence of human apolipoprotein B-100.";

RT Nucleic Acids Res. 14:7501-7503(1986).

RN (2)

RP SEQUENCE FROM N.A.

RX MEDLINE=88003974; PubMed=3652907; Ludwig E.H., Blackhart B.D., Pierotti V.R., Calaci L., Fortier C., Knott T., Scott J., Mahley R.W., Levy-Wilson B., McCarthy B.J.; "DNA sequence of the human apolipoprotein B gene.";

RT DNA 6:363-372(1987).

RN (3)

RP SEQUENCE FROM N.A.

RX MEDLINE=87008488; PubMed=3759943; Chen S.-H., Yang C.-Y., Chen P.-F., Setzer D., Tanimura M., Li W.-H., Goto A.M., Jr., Chan L.; "The complete cDNA and amino acid sequence of human apolipoprotein B-100.";

RT J. Biol. Chem. 261:12918-12921(1986).

RN (4)

RP SEQUENCE FROM N.A.

RX MEDLINE=87041416; PubMed=3464946; Law S.W., Grant S.M., Higuchi K., Hospattankar A.V., Lackner K.J., Lee N., Brewer H.B., Jr.; "Human liver apolipoprotein B-100 cDNA: complete nucleic acid and derived amino acid sequence.";

RT Proc. Natl. Acad. Sci. U.S.A. 83:8142-8146(1986).

RN (5)

RP SEQUENCE FROM N.A.

RX MEDLINE=87161758; PubMed=3030729; Cladars C., Hadzopoulou-Cladars M., Nolte R.T., Atkinson D., Zannis V.I.; "The complete sequence and structural analysis of human apolipoprotein B-100: relationship between apob-100 and apob-48 forms.";

RT EMBO J. 5:3495-3507(1986).

RN (6)

RP SEQUENCE OF 709-906 FROM N.A.

RX MEDLINE=85270450; PubMed=3860836; Deeb S.S., Motulsky A.G., Albers J.J.;

RT "A partial cDNA clone for human apolipoprotein B.";

RL Proc. Natl. Acad. Sci. U.S.A. 82:4983-4986(1985).

RN (7)

RP SEQUENCE OF 3056-3159 FROM N.A.

RX MEDLINE=86041888; PubMed=3903660; Mehrbadian M., Schumaker V.N., Fareed G.C., West R., Johnson D.F., Kirchgessner T.G., Lin H.-C., Wang X., Ma Y., Mendiaz E., Lusis A.J.; "Human apolipoprotein B: identification of cDNA clones and characterization of mRNA.";

RT Nucleic Acids Res. 13:6937-6953(1985).

RN (8)

RP SEQUENCE OF 1937-2018 AND 3811-4334 FROM N.A.

RX MEDLINE=86093680; PubMed=3841204; Carlsson P., Olofsson S.O., Bondjers G., Darnfors C., Wiklund O., Bjursell G.; "Molecular cloning of human apolipoprotein B cDNA.";

RT Nucleic Acids Res. 13:8813-8826(1985).

RN (9)

RP SEQUENCE OF 3109-4563 FROM N.A.

RX MEDLINE=85300528; PubMed=2994225; Knott T.J., Rall S.C. Jr., Innerarity T.L., Jacobson S.F., Uda M.S., Levy-Wilson B., Powell L.M., Pease R.J., Eddy R., Nakai H., Byers M., Priestley L.M., Robertson E., Rall L.B., Bethesda C., Shows T.B., Mahley R.W., Scott J.; "Human apolipoprotein B: structure of carboxyl-terminal domains, sites of gene expression, and chromosomal localization.";

RT Science 230:37-43(1985).

RN (10)

RP SEQUENCE OF 1-291 FROM N.A.

RX MEDLINE=86149325; PubMed=3513177; Procter A.A., Hardman D.A., Schilling J.W., Miller J., Appleby V., Chen G.C., Kirsner S.W., McKenzie G., Kane J.P.; "Isolation of a cDNA clone encoding the amino-terminal region of human apolipoprotein B.";

RT Proc. Natl. Acad. Sci. U.S.A. 83:1467-1471(1986).

RN (11)

RP SEQUENCE OF 1-1670 FROM N.A.

RX MEDLINE=86287319; PubMed=3461454; Procter A.A., Hardman D.A., Sato K.Y., Schilling J.W., Yamanka M., Hort Y.J., Hjertild K.A., Chen G.C., Kane J.P.; "Analysis of cDNA clones encoding the entire B-26 region of human apolipoprotein B.";

RT Proc. Natl. Acad. Sci. U.S.A. 83:5678-5682(1986).

RN (12)

RP PARTIAL SEQUENCE AND IDENTIFICATION (APO-B48).

RX MEDLINE=88018019; PubMed=3659919; Chen S.-H., Habib G., Yang C.-H., Gu Z.-W., Lee B.R., Weng S.-H., Silberman S.R., Gal S.-J., Deslypere J.P., Rosseneu M., Goto A.M., Jr., Li W.-H., Chan L.; "Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon.";

RT Science 238:365-366(1987).

RN (13)

RP DOMAINS.

RX MEDLINE=87039351; PubMed=3773997; Knott T.C., Pease R.J., Powell L.M., Wallis S.C., Rall S.C. Jr., Innerarity T.L., Blackhart B., Taylor W.R., Marcel Y., Milne R., Johnson D., Fuller M., Lusis A.J., McCarthy B.J., Mahley R.W., Levy-Wilson B., Scott J.; "Complete protein sequence and identification of structural domains of human apolipoprotein B.";

RT Nature 323:734-738(1986).

RN (14)

RP DOMAINS.

RA Yang C.-Y., Chen S.-H., Gianfrancesco S.H., Bradley W.A., Sparrow J.T., Tanimura M., Li W.-H., Sparrow D.A., Deloof H., Rosseneu M., Lee F.-S., Gu Z.-W., Goto A.M., Jr., Chan L., Rosseneu M., "Sequence, structure, receptor-binding domains and internal repeats of human apolipoprotein B-100.";

RT Nature 323:738-742(1986).

RN (15)

RP CALCIUM-BINDING DATA.

RX MEDLINE=86242245; PubMed=3087360;

RA Dahti N., Lee D.M., Mok T.;  
RT "Apolipoprotein B is a calcium binding protein.";  
RL Biochem. Biophys. Res. Commun. 137:493-499(1986).  
RN [16]  
RP VARIANT SER-4338.  
RX MEDLINE=91071750; PubMed=1979313;  
RA Navajas M., Laurent A.-M., Morel J.-F., Ragab A., Cambou J.-P.,  
RA Cuny G., Cambien F., Roizes G.;  
RT "Detection by denaturing gradient gel electrophoresis of a new  
RT polymorphism in the apolipoprotein B gene.";  
RL Hum. Genet. 86:91-93(1990).  
RN [17]  
RP VARIANT FDB GLN-3527.  
RX MEDLINE=90098975; PubMed=25631166;  
RA Soria L.F., Ludwig E.H., Clarke H.R.G., Vega G.L., Grundy S.M.,  
RA McCarthy B.J.;  
RT "Association between a specific apolipoprotein B mutation and  
RT familial defective apolipoprotein B-100.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:587-591(1989).  
RN [18]  
RP VARIANT LEU-2739.  
RX MEDLINE=91016974; PubMed=2216805;  
RA Huang L.-S., Gavish D., Breslow J.L.;  
RT "Sequence polymorphism in the human apoB gene at position 8344.";  
RL Nucleic Acids Res. 18:5922-5922(1990).  
RN [19]  
RP VARIANT FDB CYS-3558.  
RX MEDLINE=95190020; PubMed=7883971;  
RA Pullinger C.R., Hennessey L.K., Chatterton J.E., Liu W., Love J.A.,  
RA Mendel C.M., Frost P.H., Malloy M.J., Schumaker V.N., Kane J.P.;  
RT "Familial ligand-defective apolipoprotein B. Identification of a new  
RT mutation that decreases LDL receptor binding affinity.";  
RL J. Clin. Invest. 95:1225-1234(1995).  
RN [20]  
RP VARIANTS L-1437; S-1914; K-2566; T-3121; A-3945; M-4128 AND T-4481.  
RX MEDLINE=97044521; PubMed=889592;  
RA Poirier O., Ricard S., Beheque I., Souriau C., Evans A.E.,  
RA Arveiler D., Marques-Vidal P., Luc G., Roizes G., Cambien F.;  
RT "Detection of new variants in the apolipoprotein B (Apo B) gene by  
RT PCR-SSCP.";  
RL Hum. Mutat. 8:282-285(1996).  
RN [21]  
RP VARIANTS FDB GLN-3527 AND CYS-3558.  
RX MEDLINE=97403938; PubMed=9259199;  
RA Rabes J.P., Varret M., Saint-Jore B., Erlich D., Jondeau G.,  
RA Krempf M., Giraudet P., Junien C., Boileau C.;  
RT "Familial ligand-defective apolipoprotein B-100: simultaneous  
RT detection of the ARG3500-->GLN and ARG3531-->CYS mutations in a  
RT French population.";  
RL Hum. Mutat. 10:160-163(1997).  
RN [22]  
RP VARIANTS S-1914; R-1923; L-2739; D-3319; T-3427; Q-3432 AND I-3921.  
RX MEDLINE=98141125; PubMed=9490296;  
RA Lerin T.P., Bakken K.S., Hoel V., Hjermann I., Berg K.;  
RT "Screening for mutations of the apolipoprotein B gene causing  
RT hypocholesterolemia.";  
RL Hum. Genet. 102:44-49(1998).  
RN [23]  
RP FUNCTION: APOLIPOPROTEIN B IS A MAJOR PROTEIN CONSTITUENT OF  
CC CHYLOMICRONS, VLDL AND LDL. IT FUNCTIONS AS A RECOGNITION SIGNAL  
CC FOR THE CELLULAR BINDING AND INTERNALIZATION OF LDL PARTICLES BY  
CC THE APOB/E RECEPTOR.  
CC -!- DISEASE: DEFECTS IN APOB ARE A CAUSE OF FAMILIAL LIGAND-DEFECTIVE  
CC APOLIPOPROTEIN B-100 (FDB). IT IS A DOMINANTLY INHERITED DISORDER  
CC OF LIPOPROTEIN METABOLISM LEADING TO HYPERCHOLESTEROLEMIA AND  
CC INCREASED PRONENESS TO CORONARY ARTERY DISEASE (CAD).  
CC THE PLASMA CHOLESTEROL LEVELS ARE DRAMATICALLY ELEVATED DUE TO  
CC IMPAIRED CLEARANCE OF LDL PARTICLES BY DEFECTIVE APOB/E RECEPTORS.  
CC -!- DISEASE: DEFECTS IN APOB ASSOCIATED WITH OTHER GENE DEFECTS  
CC (POLYGENIC) CAN BE THE CAUSE OF HYPOCHOLESTEROLEMIA.  
CC -!- MISCELLANEOUS: APO B-48 WHICH IS SYNTHESIZED ONLY BY THE  
CC INTESTINE, AND FOUND IN CHYLOMICRONS, IS A SHORTENED FORM OF APO

Query Match

4.9%; Score 92; DB 1; Length 4563;

Best Local Similarity 20.7%; Pred. No. 99;  
Matches 61; Conservative 48; Mismatches 118; Indels 68; Gaps 12;  
QY 3 YCERLDPKYLKIALKRLQSLSPFQNSQSLQRAYSTPYSYR---IILQKENKQAL 59  
DB 4211 YTREELCTMFIREVGTVLVSQVSKVHNGSEIL-----FSYFQDLVITLPFE-----L 4257  
QY 60 ARHKCISITLFEFFKNLLFVHLLSLSKNOREGCTDMVSTPFFNRNLW-----YRLSS 113  
DB 4258 RKHLKIDVISMYREL---LKLSKEAQEVFKAIOSLKTKTEVL-RNLQDLLQPIFQIED 4312  
QY 114 RFSLWMSYCPFFFLDYL--EAGLLSDFLDHOAVIKPFLETHSFYSYYPVSGFVAP----- 166  
DB 4313 NIKQLKEMKFTYLYNIQDEINTIFNDYIPY--VFLLKENLCLNLHKFNEFQNELQEA 4370  
QY 167 -----HOYLSILODRYFFIASVMT-----LDKDNESLTPDLI----- 199  
DB 4371 SOELQOIHQYIMALREYFDPISIVGWTVKYVELEKIVSLKNLVALKDFHSEYIVSAS 4430  
QY 200 ---HDLGHVPWLHPSEFFFNMGRLFTKVIKQALPSPKKORIOTLOSNLIA 251  
DB 4431 NFTSOLSSQVEQFLHRNIQIYLSILTPDPKGKKEIAELSAQAQEI--IKSQAIA 4483  
RESULT 48  
ID SPEE METJA STANDARD; PRT; 293 AA.  
AC Q57761;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Probable spermidine synthase (EC 2.5.1.16) (Putrescine  
DE aminopropyltransferase) (SPDSY).  
GN SPEE OR M30313.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
OX NCBI\_TaxID=2190;  
RN [1]  
RP SEQUENCE FROM N.A. 2661 / ATCC 43067;  
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=8689087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
RT jannaschii.";  
RL Science 273:1058-1073(1996).  
CC -!- CATALYTIC ACTIVITY: S-adenosylmethionine + putrescine = 5'-  
CC methylthioadenosine + spermidine.  
CC -!- PATHWAY: FIFTH (LAST) STEP IN THE BIOSYNTHESIS OF SPERMIDINE FROM  
CC ARGinine AND METHIONINE.  
CC -!- SIMILARITY: BELONGS TO THE SPERMIDINE/SPERMINE SYNTHASES FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U67486; AAB98300.1; -.  
CC TIGR; M30313; -.  
CC InterPro; IPR000051; SAM bind.  
CC InterPro; IPR001045; Spermine synthase.  
CC Pfam; PF01564; Spermine synth; 1.  
CC TIGRFAMs; TIGR00417; speE; 1.  
CC -----

DR PROSITE: P801330, SPERMIDINE SYNTHASE; 1.  
 KM Spermidine biosynthesis; transferase; Complete proteome.  
 FT DOMAIN 85 126 BINDING TO DECARBOXYLATED SAM  
 (POTENTIAL)  
 SQ SEQUENCE 293 AA; 33899 MW; 88848C919E5452A CRC64;  
 Query Match 4.8%; Score 91.5; DB 1; Length 293;  
 Best Local Similarity 19.1%; Pred. No. 3.4;  
 Matches 46; Conservative 36; Mismatches 72; Indels 87; Gaps 11;  
 OY 157 YPVSGFVAPHOYLSLQDRYFPIASVM---RTLDKDNFSITPDLIHDLGHVFWLHP 212  
 DB 32 YREKSGF---QEIIIDTYDFGKALILDNTFQTERDEF---IYHLLSHIPLFTHP 82  
 OY 213 SFSEFFINMG-----RFTKVIKVOA-LPSKKORIQTLSNLI 250  
 DB 83 NRRNVLVIQGGGGTREVVGKSVETVDFVELDEKVIKACKKMPKLSCEIDNEKVL 142  
 OY 251 AIVRCFWTVSSGLIENHGRKAYGAVLISSPOLGHAFLDNVRVLPFLDQIRLPNT 310  
 DB 143 -----INDGIKVAETEKKDVIIVDCPPVG----- 169  
 OY 311 STPOETLSIRHFDLVLTSLKLEMLDQGL--ESIPLYOE-----KYL--SGREV 359  
 DB 170 --PAKGLFEKPEYKAVFKCLN-----DDGIWQSESPFLNLDLQNICRYLADGFKI 221  
 OY 360 L 360  
 DB 222 I 222  
 RESULT 49  
 RRLP\_MEASA STANDARD; PRT; 2183 AA.  
 ID RRLP\_MEASA  
 AC P35975;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE RNA polymerase beta subunit (EC 2.7.7.48) (Large structural protein)  
 DE (L. protein).  
 GN L.  
 OS Measles virus (strain Aik-C) (Subacute sclerosing panencephalitis  
 OS virus); sRNA negative-strand viruses; Mononegavirales;  
 OC Paramyxoviridae; Paramyxovirinae; Morbilliviruses.  
 OX NCBI\_TaxID=36408;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93227570; PubMed=8470368;  
 RA Mori T., Sasaki K., Hashimoto H., Makino S.;  
 RT "Molecular cloning and complete nucleotide sequence of genomic RNA of  
 RT the Aik-C strain of attenuated measles virus.";  
 RL Virus Genes 7:67-81(1993).  
 CC -1- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY  
 CC SYNTHESIZE IN RNA SYNTHESIS, CAPING, METHYLATION AND POLY(A)  
 CC SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAs, RNA EDITING OF THE P  
 CC GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC {RNA} (N).  
 CC -1- SIMILARITY: WITH L. PROTEIN OF OTHER PARAMYXOVIRUSES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by, and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC EMBL: S58435; AAB26147.1; --  
 DR PIR: G48556; G48556.  
 DR InterPro: IPR001016; Viral\_RNA\_pol\_L.  
 DR Pfam: PF00946; Paramyx\_RNA\_pol; 1.

KM Transferase; RNA-directed RNA polymerase.  
 SQ SEQUENCE 2183 AA; 247754 MW; 12A6868F4C8A2199 CRC64;  
 Query Match 4.8%; Score 91; DB 1; Length 2183;  
 Best Local Similarity 20.6%; Pred. No. 47;  
 Matches 78; Conservative 46; Mismatches 119; Indels 136; Gaps 18;  
 OY 57 QALAHKCSITLFFKNLLFVLLSLSKNREGSGTMAVAVSPFFNRMLWRLSSRPS 116  
 DB 429 EGLTHREQVDNMKSFAGVF-----GCFMPLSLSDSD--LTMVLLKKAALAAQR 474  
 OY 117 LMKSCPRFLDYLEAFGLSDFLDQAVIKFELFTHFSYVSGFVAPHOYLSLQDR 176  
 DB 475 EMDSVYKPEFLRDYDPKGTGS---RRLVDFVFNDSFPDYVIMVYSGAV---LHDP 526  
 OY 177 YFPIASVMRTLDKDNFSITPDLIHDLGHVFWLHPFSBFFINMGRLEFTKVIKVOALP 236  
 DB 527 EFNL-----SYLKEKEIK-----TGRLEFAKMTYKRA-- 555  
 OY 237 SKKORIQTLSNLI-----IVRCFWTVSSG-----LIENHGRKAY 274  
 DB 556 -----GVINENLISNGIKYFKDNGMAKDEHDTYKALHLAVSGVPKDKESHRO-- 606  
 OY 275 GAVLISSPOLGHAFLDNVRVLP--LELDQIRLPNTSTPOE----- 315  
 DB 607 GPVLKTYSRSPVHTSTRNVRRAKGFIPGPYIRODDTHPENMEAEVTSAPITTDLKK 666  
 OY 316 -----TLESIRHFDLVLTSLKLEML--LDQGLL-----ESIPLY-- 348  
 DB 667 YCLNMRYETISLPAQR-LMEIYGLPSFPQWLHKRLFTSVLVSDPHCPDPLDAHIFLYKV 725  
 OY 349 -NOE---KY-LSGFVTLQ 362  
 DB 726 PNDQIFIKYPMGIGIEGVCQ 744  
 RESULT 50  
 RRLP\_MEASA STANDARD; PRT; 2183 AA.  
 ID RRLP\_MEASA  
 AC P32576;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE RNA polymerase beta subunit (EC 2.7.7.48) (Large structural protein)  
 DE (L. protein).  
 GN L.  
 OS Measles virus (strain Edmonston) (Subacute sclerosing panencephalitis  
 OS virus).  
 OC Paramyxoviridae; Paramyxovirinae; Morbilliviruses;  
 OC Viruses; sRNA negative-strand viruses; Mononegavirales;  
 OX NCBI\_TaxID=11235;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88219537; PubMed=2835864;  
 RA Blumberg B.M., Crowley J.C., Silverman J.I., Menonna J., Cook S.D.,  
 RA Dowling P.C.;  
 RT "Measles virus L protein evidences elements of ancestral RNA  
 RT polymerase.";  
 RL Virology 164:487-497(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90085790; PubMed=2596022;  
 RA Cattaneo R., Schmid A., Spielhofer P., Kaelin K., Baccaro K.,  
 RA Melien V., Padowitz J., Flanagan S., Rima B.K., Udem S.A.;  
 RT "Mutated and hypermutated genes of persistent measles viruses which  
 RT caused lethal human brain diseases.";  
 RL Virology 173:415-425(1989).  
 CC -1- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY  
 CC SYNTHESIZE IN RNA SYNTHESIS, CAPING, METHYLATION AND POLY(A)  
 CC SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAs, RNA EDITING OF THE P  
 CC GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC {RNA} (N).



Fri Jan 10 10:56:13 2003

```
CC -!- SIMILARITY: WITH L PROTEIN OF OTHER PARAMYXOVIRUSES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M20865; AAA46430.1; -
CC DR EMBL; K01711; AAA75501.1; -
CC DR PIR; A28919; ZLNZMV.
CC DR InterPro; IPR001016; Viral_RNA_pol_L.
CC DR Pfam; PF00946; Paramyx_RNA_pol; 1.
CC DR Transferrase; RNA-directed RNA polymerase.
CC SQ SEQUENCE 2183 AA; 247646 MW; 1B0B03CA2E2B6EA5 CRC64;
CC -----
Query Match 4.8%; Score 91; DB 1; Length 2183;
Best Local Similarity 20.6%; Pred. No. 47;
Matches 78; Conservative 46; Mismatches 119; Indels 136; Gaps 18;
QY 57 QALARHKCISILEEFKNLLFVHLLSLSKNQREGCSTDMVVSTPFFNRLMYRLSLSRFS 116
Db 429 EGLTHEQCVDNWKSFAGKP-----GCFWPLSLSD--LTWYLDKALAAALQR 474
QY 117 LMKSYCPFFLDYLEAFGLLSDFLDHQAVIKFELETHPSYYPVSGFVAPHQYLSLQDR 176
Db 475 EWDSVYPKEFLRYDPKGTGS-----RRLDVFLNDSSFDPYDVMYVSGAY---LHDP 526
QY 177 YFPIASVMRTLDKDNFSLTDPDLHDLGHVPWMLHPSFSEFFINNGRLFTKVIKVOALP 236
Db 527 EENL-----SYSLQKEIKE-----TGRLFAMTYKMR-- 555
QY 237 SKQRIQTLQSNLIA-----IVRCFWFTVESG-----LIENHEGRKAY 274
Db 556 -----CQVIAENLISNGIGKYGKNGMAKDEODLTALHTLAVSGVPKDLKESHRG--- 606
QY 275 GAVLISSPOELGHAFIDNVRLP--LELDQIRLPNTSTPOE----- 315
Db 607 GPVLKTYSRSPVHTSTRVRAAKGFIGFPQVIRQODTDHPENMEAVETVSAFITTDLKK 666
QY 316 -----TLFSIRHFDLVELTSKLEWM--LDQGLL-----ESIPLY-- 348
Db 667 YCLNWRYETISLFAQR-LNEYGLPSPFQWLHKRLTSLVLYVSDPHCPDLDLDAHPLYKV 725
QY 349 -NQE---KY-LSGFEVLCO 362
Db 726 PNDQIFIKYPMGGIEGYCQ 744
```

Search completed: January 9, 2003, 15:27:50  
Job time : 33 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2003, 11:51:03 ; Search time 65 Seconds  
(without alignments)

742.103 Million cell updates/sec

Title: US-09-438-185A-1047

Perfect score: 1889

Sequence: 1 VHYCETLDPKYLKIALKL.....ESTPLNQKYLGSFVLQ 362

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : 1\_ /SIDS2/gcgdata/geneseq/genesecp-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/genesecp-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/genesecp-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/genesecp-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/genesecp-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/genesecp-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/genesecp-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/genesecp-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/genesecp-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/genesecp-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/genesecp-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/genesecp-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/genesecp-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/genesecp-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/genesecp-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/genesecp-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/genesecp-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/genesecp-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/genesecp-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/genesecp-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/genesecp-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/genesecp-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/genesecp-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1889	100.0	362	23	AB90626
2	1250.5	66.2	259	20	AA935703
3	237	12.5	429	22	AAE11086
4	235.5	12.5	415	22	AAE11161
5	231.5	12.3	444	22	AA80192
6	227.5	12.0	444	22	AA80192
7	226.5	12.0	434	23	AAO17329
8	222	11.8	498	15	AA950185
9	221	11.7	498	14	AA936741
10	220	11.6	497	15	AA950184

Human tyrosine hyd  
Drosophila melanog  
Phenylalanine hydr  
Human phenylalanin  
Human phenylalanin  
Human phenylalanin  
Caenorhabditis ele  
Novel human secret  
Drosophila melanog  
Drosophila melanog  
Novel human secret  
phAB fusion pepti  
Human ORFX protein  
Drosophila melanog  
Drosophila melanog  
Fungal ZAC protein  
Human protein sequ  
Drosophila melanog  
Amino acid sequenc  
Mycoplasma genital  
phAB fusion pepti  
Novel human protei  
C. pneumoniae prot  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
H. pylori ORF hp5p  
S. epidermidis ope  
Yeast transcriptio  
Lactococcus lactis  
Novel human diagn  
RhCMV unique short  
Novel human secret  
Human Apo B protei  
Apollipoprotein B-1  
Amino acid sequenc  
Thermoanaerobacter  
Thermoanaerobacter  
T. thermohydrosulf  
Drosophila melanog  
Drosophila melanog  
Drosophila melanog  
Measles virus Rube  
Measles virus Mora  
Measles virus Zagz  
Measles virus Aik-  
Measles virus Edmo  
AmEPV subunit of ye  
ORC3 subunit of ye  
S. cerevisiae orig  
Staphylococcus epi  
Staphylococcus epi  
Amino acid sequenc  
L protein of atten  
Novel human secret  
Novel human diagn  
Pseudomonas Orxys  
Pseudomonas alcali  
Pseudomonas alcali  
Pseudomonas alcali  
Drosophila melanog  
Human olfactory re  
Drosophila melanog  
Drosophila melanog  
Novel human diagn  
Novel human diagn  
Human breast cance  
Protein encoded by  
Amino acid sequenc  
Drosophila melanog  
Amino acid sequenc  
Splice variant of  
Splice variant of

84	88	4.7	763	22	AAU18676	Renal and cardio
85	87.5	4.6	297	22	AA65297	Human interleukin-
86	87.5	4.6	309	20	AA39785	Interleukin-18 rec
87	87.5	4.6	328	20	AA39789	Interleukin-18 rec
88	87.5	4.6	521	20	AA39780	Interleukin-18 rec
89	87.5	4.6	540	19	AAW7745	Human interleukin-
90	87.5	4.6	540	20	AAJ39786	Interleukin-18 rec
91	87.5	4.6	541	18	AAJ31207	Human receptor pro
92	87.5	4.6	541	20	AAI06513	Human interleukin-
93	87.5	4.6	541	23	AAI14731	Human IL-18 recept
94	87.5	4.6	541	23	AAI16583	Human interleukin-
95	87.5	4.6	1817	21	AAI18255	Plasmodium falcipa
96	87	4.6	293	21	AAV94384	Helicobacter pylor
97	87	4.6	939	21	AA553192	Maccaca mulatta rha
98	86.5	4.6	199	23	ABP28782	Streptococcus poly
99	86.5	4.6	210	22	AAW90773	Human immune/haema
100	86.5	4.6	244	23	ABP30079	Streptococcus poly
101	86.5	4.6	571	22	ABG09119	Novel human diagno
102	86.5	4.6	571	22	ABG18446	Novel human diagno
103	86.5	4.6	990	14	AA39343	Epib protein. Sta
104	86.5	4.6	990	14	AA43431	S. epidermis readi
105	86	4.6	656	23	ABP39477	Staphylococcus epi
106	86	4.6	662	22	AB62836	Drosophila melanog
107	86	4.6	924	22	AAE04772	Human vesicle traf
108	85.5	4.5	486	22	AAW93848	Human polypeptide,
109	85.5	4.5	799	20	AAW86352	Human DNAX toll-1i
110	85.5	4.5	799	23	AAH83162	Human Toll-1-like re
111	85.5	4.5	799	23	AAE16093	Human DNAX Toll-1i
112	85.5	4.5	837	20	AAE86361	Human DNAX Toll-1i
113	85.5	4.5	837	23	AAE16102	Human DNAX Toll-1i
114	85.5	4.5	837	23	AAE16116	Human DNAX Toll-1i
115	85	4.5	471	21	AAI18201	Plasmodium falcipa
116	85	4.5	549	19	AAW98659	H. pylori GHPo 467
117	85	4.5	774	22	AA393433	Human protein sequ
118	85	4.5	924	22	AAW40111	Human polypeptide
119	85	4.5	959	22	ABG02299	Novel human diagno
120	85	4.5	2183	19	AAW48704	Measles virus 1977
121	84.5	4.5	520	22	AB611995	Drosophila melanog
122	84.5	4.5	867	22	AAE00434	Quail sulfatase (Q
123	84.5	4.5	871	21	AAE00191	Breast cancer prot
124	84.5	4.5	871	22	AAE85481	Human 23553 sulfat
125	84.5	4.5	1091	22	ABG14938	Novel human diagno
126	84	4.4	314	23	AB684755	DNA polymerase III
127	84	4.4	324	14	AA33388	BSH from Lactobaci
128	84	4.4	978	22	AAW58302	Human brain expres
129	84	4.4	2785	21	AAV57148	Human down-regulat
130	83.5	4.4	317	22	AAW96470	Purified P. abyssi
131	83.5	4.4	450	22	AAW98337	Escherichia coli p
132	83.5	4.4	835	22	AAU07869	Polypeptide sequen
133	83.5	4.4	891	22	ABG25721	Novel human diagno
134	83.5	4.4	1107	22	ABG18247	Novel human diagno
135	83	4.4	358	23	AB649302	Listeria monocytog
136	83	4.4	405	22	AAW96288	Purified P. abyssi
137	83	4.4	491	20	AAW16133	A murine Rint. poly
138	83	4.4	539	22	ABG06512	Novel human diagno
139	83	4.4	539	22	ABG26540	Novel human diagno
140	83	4.4	3423	22	ABG23417	Novel human diagno
141	82.5	4.4	463	22	ABG70150	Drosophila melanog
142	82.5	4.4	956	23	AAW87729	Human calyculenin-
143	82.5	4.4	956	23	AAW06223	Peptide #4905 enco
144	82.5	4.4	1081	20	AAV13443	Peptide #4905 enco
145	82.5	4.4	1105	20	AAV13442	Amino acid sequenc
146	82.5	4.4	1202	22	ABW31470	Peptide #4121 enco
147	82.5	4.4	1202	22	ABW31470	Peptide #4121 enco
148	82.5	4.4	1202	22	AAW57442	Human brain expres
149	82.5	4.4	1202	22	AAW17663	Peptide #4097 enco
150	82.5	4.4	1270	22	ABG06368	Novel human diagno

## ALIGNMENTS

ABW90626	ABW90626 standard; Protein; 362 AA.
ID	ABW90626 standard; Protein; 362 AA.
XX	
AC	ABW90626;
XX	
XX	
DT	29-JUL-2002 (first entry)
XX	
DE	Chlamydia pneumoniae CP7380 protein, SEQ ID NO:201.
XX	
XX	
KW	Chlamydial infection; antigen; immunogen; vaccine; diagnosis;
KW	human respiratory disease; cardiovascular disease; atherosclerosis;
KW	coronary artery disease; carotid artery stenosis; myocardial infarction;
KW	cerebrovascular disease; aortic aneurysm; claudication; stroke;
KW	strain CML029.
OS	Chlamydia pneumoniae.
XX	
XX	WO200202606-A2.
PN	
XX	10-JAN-2002.
PD	
XX	
XX	03-JUL-2001; 2001WO-1B01445.
PF	
XX	
XX	03-JUL-2000; 2000GB-0016363.
PR	11-JUL-2000; 2000GB-0017047.
PR	21-JUL-2000; 2000GB-0017983.
PR	07-AUG-2000; 2000GB-0019368.
PR	18-AUG-2000; 2000GB-0020440.
PR	14-SEP-2000; 2000GB-0022583.
PR	10-NOV-2000; 2000GB-0021542.
PR	22-DEC-2000; 2000GB-0031706.
PA	(CHIR-) CHIRON SPA.
XX	
PI	Ratti G, Grandi G;
XX	
XX	WPI: 2002-154726/20.
DR	N-PSDB; ABL91284.
XX	
PT	Novel Chlamydia pneumoniae protein useful in the manufacture of a
PT	medicament for treatment or prevention of infection due to Chlamydia,
PT	preferably Chlamydia pneumoniae, and for diagnostic purposes -
PS	Claim 1; Page 134; 364pp; English.
XX	
XX	Sequences ABW90526-ABW90715 represent novel proteins from Chlamydia
CC	pneumoniae (strain CML029), and ABL91184-ABL91373 represent DNA encoding
CC	them. The proteins are predicted to be immunogenic and may therefore be
CC	useful in vaccine production and for diagnostic purposes. Chlamydia
CC	pneumoniae is a common cause of respiratory disease in humans, and is
CC	also involved in the development of cardiovascular diseases such as
CC	atherosclerosis, coronary artery disease, carotid artery stenosis,
CC	myocardial infarction, cerebrovascular disease, aortic aneurysm,
CC	claudication and stroke. The proteins and nucleic acids of the invention
CC	may be used in vaccines and pharmaceutical compositions for the
CC	prevention or treatment of chlamydial infections, particularly Chlamydia
CC	pneumoniae infections. The proteins may also be used in the detection of
CC	Chlamydia pneumoniae, and the nucleic acids may be used in PCR, branched
CC	DNA probe assay or blotting techniques for determining Chlamydia
CC	pneumoniae gene expression. The present sequence represents a
CC	specifically claimed Chlamydia pneumoniae protein of the invention.
XX	
SO	Sequence 362 AA;
Query March	100.0%; Score 1889; DB 23; Length 362;
Best Local Similarity	100.0%; Pred. No. 8; 6e-182;
Matches 362; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 VHCERTLPDKYTLTKATKROSLSLAFRONSOSLORAYSPRYVYRIILQENKKEKQALA 60
DB	1 VHCERTLPDKYTLTKATKROSLSLAFRONSOSLORAYSPRYVYRIILQENKKEKQALA 60
QY	61 RHKCIStLEPFKNLLFVHLILSLSKNQRGCSTDMANVSTPEPNRNLWYRLLSRFLWKS 120

Db 61 RHKGISILEFFKNLLFVHLLSLSKNQREGSCDMAVSTPPFNRLWYRLSSRFLWKS 120  
QY 121 YCPREFLDYLEAFGLLSDFLDHQAVIKFPELETHFSYYPVSGFVAPHOYLSLLQDRYPP 180  
Db 121 YCPREFLDYLEAFGLLSDFLDHQAVIKFPELETHFSYYPVSGFVAPHOYLSLLQDRYPP 180  
QY 181 ASVVRTLDKDNFSLTPDLIHDLGHVPWLLHPSFSEFFINNGRLFTKVIKVOALPSKKQ 240  
Db 181 ASVVRTLDKDNFSLTPDLIHDLGHVPWLLHPSFSEFFINNGRLFTKVIKVOALPSKKQ 240  
QY 241 RIOTLQSNLAIVRCPFVTFVESGLIENHGRKAYGAVLISSPQELGHAFIDNVRLPLEL 300  
Db 241 RIOTLQSNLAIVRCPFVTFVESGLIENHGRKAYGAVLISSPQELGHAFIDNVRLPLEL 300  
QY 301 DOIIRLPNTSTPOETLFSIRHPDELVELTSKLEWMLDQGLLESIPLYNQEKYLSGFEVL 360  
Db 301 DOIIRLPNTSTPOETLFSIRHPDELVELTSKLEWMLDQGLLESIPLYNQEKYLSGFEVL 360  
QY 361 CQ 362  
Db 361 CQ 362

RESULT 2  
AAV35703  
ID AAV35703 standard; Protein; 259 AA.  
AC AAV35703;  
DT 13-SEP-1999 (first entry)  
DE Chlamydia pneumoniae transmembrane protein sequence.  
XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
KW vaccine; neutralising epitope.  
OS Chlamydia pneumoniae.  
XX WO9927105-A2.  
XX 03-JUN-1999.  
XX 20-NOV-1998; 98WO-IB01890.  
XX 04-NOV-1998; 98US-0107078.  
XX 21-NOV-1997; 97FR-0014673.  
XX (G8ST ) GENSET.  
XX Griffiths R;  
XX WPI; 1999-357842/30.  
XX Genome sequence of Chlamydia pneumoniae  
XX Page 1410-1411; Disclosure; 1912pp; English.  
XX AAY34594-Y35879 represent the proteins encoded by all the open reading  
CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.  
CC C. pneumoniae causes respiratory disease such as pneumonia and  
CC bronchitis and is thought to be a contributing factor in heart  
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in  
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
CC nucleotide sequences can also be used as immunogenic compositions,  
CC especially where the vector directs the expression of a neutralising  
CC epitope of C. pneumoniae.  
XX Sequence 259 AA;  
SQ

Query Match 66.2%; Score 1250.5; DB 20; Length 259;  
Best Local Similarity 91.8%; Pred. No. 1.3e-117;  
Matches 245; Conservative 3; Mismatches 8; Indels 11; Gaps 1;  
QY 96 VVSTPPFNRLWYRLSSRFLWKSVCPRFDFLDYLEAFGLLSDFLDHQAVIKFPELETHF 155  
Db 4 MVSTPFLTVFSMEKLLS-----KIFLDYLEAFGLLSDFLDHQAVIKFPELETHF 52  
QY 156 SYYPVSGFVAPHOYLSLLQDRYFPPTASVMRTLDKDNFSLTPDLIHDLGHVPWLLHPSFS 215  
Db 53 SYYPVSGFVAPHOYLSLLQDRYFPPTASVMRTLDKDNFSLTPDLIHDLGHVPWLLHPSFS 112  
QY 216 EFFINMGRLFTKVIKVOALPSKKQRIOTLQSNLAIVRCPFVTFVESGLIENHGRKAYG 275  
Db 113 EFFINMGRLFTKVIKVOALPSKKQRIOTLQSNLAIVRCPFVTFVESGLIENHGRKAYG 172  
QY 276 AVLSSPQELGHAFIDNVRLPLELDOIIRLPNTSTPOETLFSIRHPDELVELTSKLEW 335  
Db 173 AVLSSPQELGHAFIDNVRLPLELDOIIRLPNTSTPOETLFSIRHPDELVELTSKLEW 232  
QY 336 MLDQGLLESIPLYNQEKYLSGFEVLQ 362  
Db 233 MLDQGLLESIPLYNQEKYLSGFEVLQ 259  
RESULT 3  
AAE11086  
ID AAE11086 standard; Protein; 429 AA.  
AC AAE11086;  
DT 18-DEC-2001 (first entry)  
DE Protein encoded by pFUSAB vector DNA insert sequence.  
XX Phenylalanine hydroxylase; PAH; phhA; 4a-carbinolamine dehydratase; phhB;  
KW aromatic aminotransferase; phhC; milk protein; animal protein; casein;  
KW proteinaceous food product; globulin; whey protein; phenylketonuria;  
KW PKU; inherited metabolic disorder; impaired brain function; nootropic;  
KW cell therapy.  
XX Chimeric - Escherichia coli.  
OS Chimeric - Unidentified.  
XX Key Location/Qualifiers  
FT Region 1..26 /label= LacZ\_peptide  
FT Region 27..28 /label= Linker\_peptide  
FT Region 29..289 /label= phhA protein  
FT Region 290..311 /label= Junction\_peptide  
FT Region 312..429 /label= phhB protein  
XX WO20016822-A2.  
XX 20-SEP-2001.  
XX 14-MAR-2001; 2001WO-DK00172.  
XX 14-MAR-2000; 2000US-0525116.  
XX (NILA-) NILAB APS.  
XX Johnsen M, Ravn P, Madsen SM, Vrang A, Israelsen H, Bredmose L;  
PI Arnau J, Jensen SH, Gjetting T, Nielsen E;  
XX WPI; 2001-590055/66.  
XX N-PSDB; AAD18533.  
XX Novel recombinant cells comprising a nucleic acid encoding a gene  
PT



QY 207 PMLHPSFSEFFINMGRLETKVIEKQALPSKQRIQTLQSNLIAIVRCFWFTVSGSLIE 266  
 Db 141 PLLTNPWFATHTYKGLKA-----SKER-----VFLARLYWNTIEFGLVE 184  
 QY 267 NHEGRKAYGAVLISSPOELGHAFID--NVRVLELDQIIRLPFTSTPQETLSIRHF 323  
 Db 185 TDQKRIYGGIILSSPRETVISLSDEPLHAFNELE---AWRTPIRIDILQLYVLPDL 241  
 QY 324 DELVELTSK 332  
 Db 242 KRUFQIAQE 250  
 RESULT 5  
 ID AAM80192  
 AC AAM80192 standard; Protein; 444 AA.  
 XX  
 XX AAM80192;  
 XX  
 XX 06-NOV-2001 (first entry)  
 XX  
 XX Human protein SEQ ID NO 3838.  
 XX  
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorder; arthritis; inflammation.  
 XX  
 XX Homo sapiens.  
 XX  
 XX WO200157190-A2.  
 XX  
 XX 09-AUG-2001.  
 XX  
 XX 05-FEB-2001; 2001WO-US04098.  
 XX  
 XX 03-FEB-2000; 2000US-0496914.  
 XX  
 XX 27-APR-2000; 2000US-0560875.  
 XX  
 XX 20-JUN-2000; 2000US-0598075.  
 XX  
 XX 19-JUL-2000; 2000US-0620325.  
 XX  
 XX 01-SEP-2000; 2000US-0654936.  
 XX  
 XX 15-SEP-2000; 2000US-0663561.  
 XX  
 XX 20-OCT-2000; 2000US-0693325.  
 XX  
 XX 30-NOV-2000; 2000US-0728422.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 XX  
 XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
 XX  
 XX WPI; 2001-476283/51.  
 XX  
 XX N-PSDB; AAKS3325.  
 XX  
 XX Nucleic acids encoding polypeptides with cytokine-like activities,  
 PT useful in diagnosis and gene therapy -  
 XX  
 XX Claim 20; Page 446-447; 6221pp; English.  
 XX  
 XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAM80020) are omitted as the relevant pages from the sequence listing

CC were missing at the time of publication.  
 SQ Sequence 444 AA;  
 Query Match 12.3%; Score 231.5; DB 22; Length 444;  
 Best Local Similarity 27.8%; Pred. No. 1.6e-14;  
 Matches 61; Conservative 43; Mismatches 96; Indels 27; Gaps 4;  
 QY 107 WYRLSSRPSLWKSYPFFLDYLEAFGLLSDFLDH-----QAVIKFPELETHFSY 158  
 Db 174 WGTVFQELNKLYPTHACR---EYLKMLLSKYCYGREDNIPQLEDVSNFLKERTGFSIR 230  
 QY 159 PVSGFVAPHOYLSLQDRYFPIASVMRTLDKDNFSLTPDLIHLGLHVPMLHPSPSEPF 218  
 Db 231 PVAGYLSPRDFLSGLAFRVFCHCTQYVRHSSDPFYTPBPDTCHELLGHVPLLAAPSPAQFS 290  
 QY 219 INMGRLETKVIEKQALPSKQRIQTLQSNLIAIVRCFWFTVSGSLIENHEGRKAYGAVL 278  
 Db 291 QEIG-----LASLGASEAVQKLAT-----CYFFTVEFLCKODQQLRVFAGL 334  
 QY 279 ISSPQELGHAFIDNVRVLELDQIIRLPFTSTPQETLSIRHFDELVE 328  
 Db 335 LSSISELKHLSGHAKVFPDKITCKOECLITTFQDVYFVSESFEDAKE 384  
 RESULT 6  
 ID AAM79208  
 AC AAM79208 standard; Protein; 444 AA.  
 XX  
 XX AAM79208;  
 XX  
 XX 06-NOV-2001 (first entry)  
 XX  
 XX Human protein SEQ ID NO 1870.  
 XX  
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorder; arthritis; inflammation.  
 XX  
 XX Homo sapiens.  
 XX  
 XX WO200157190-A2.  
 XX  
 XX 09-AUG-2001.  
 XX  
 XX 05-FEB-2001; 2001WO-US04098.  
 XX  
 XX 03-FEB-2000; 2000US-0496914.  
 XX  
 XX 27-APR-2000; 2000US-0560875.  
 XX  
 XX 20-JUN-2000; 2000US-0598075.  
 XX  
 XX 19-JUL-2000; 2000US-0620325.  
 XX  
 XX 01-SEP-2000; 2000US-0654936.  
 XX  
 XX 15-SEP-2000; 2000US-0663561.  
 XX  
 XX 20-OCT-2000; 2000US-0693325.  
 XX  
 XX 30-NOV-2000; 2000US-0728422.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 XX  
 XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
 XX  
 XX WPI; 2001-476283/51.  
 XX  
 XX N-PSDB; AAK52341.  
 XX  
 XX Nucleic acids encoding polypeptides with cytokine-like activities,  
 PT useful in diagnosis and gene therapy -  
 XX  
 XX Claim 20; Page 4255-4256; 6221pp; English.  
 XX  
 XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity relating to

CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoietic regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAM80020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.

CC Sequence 444 AA;

Query Match 12.0%; Score 227.5; DB 22; Length 444;  
 Best Local Similarity 27.4%; Pred. No. 4.2e-14;  
 Matches 63; Conservative 43; Mismatches 97; Indels 27; Gaps 4;

OY 107 WYRLSSRFLSMKSCYCPREFLDYLAFLGLSDFLDH-----QAVIKFELETHFSY 158  
 DB 174 WGTFOELNKLYPTHACK---EYLKMLPLSKICGIRENITPLEDVSNFKERTGFSIR 230  
 OY 159 PVSGFVAAPHOYLSTLDQRYPIASVMRTLDKDNFSLTPDLIHLLGHVPMILHPSSEFF 218  
 DB 221 PVAGYLSRDFLSGLAFRFVHCTQYVRHSSDPYTPPEPTCHELGHVPLAEPSPAFQS 230  
 OY 219 INMGRLEFTKVIKQVALPSKKORIQTLOSNIATVRCFWFTVESGLIENHGRKAYGAVL 278  
 DB 291 OEIG-----LASLGASEEAVQKLTAT-----CYFFTVFGLCKQDQGLRVFGAGL 324  
 OY 279 ISSPOELGHAFIDNVRVLPLELDQIRLPFNTSPQETLSIRHDELVE 328  
 DB 335 LSSISELKHVLSGHAKVFPDPKITCKQECILTTFQDVYFVESPEDAKE 384

#### RESULT 7

ID AAO17329 standard; Protein; 434 AA.

AC AAO17329;

DT 08-JUL-2002 (first entry)

DE Neurone specific tryptophane hydroxylase.

KW Neuronal tryptophane hydroxylase; TPH; neuronal disease;

KW primary haemostasis deficiency; allergy; transplantation; serotonin;

KW antiarteriosclerosis; thrombolytic; anticoagulant; immunosuppressive;

KW antiallergic; gynaecological.

OS Unidentified.

PN WO200217891-A2.

PD 07-MAR-2002.

PF 27-AUG-2001; 2001WO-DE03178.

PR 31-AUG-2000; 2000DE-1043124.

PA (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.

PI Walther D. Bader M;

DR WPI; 2002-281015/32.

PT Regulating levels of serotonin, useful for diagnosing neuronal disease

PT and treating primary hemostatic deficiency, comprises controlling

PT activity of tryptophan hydroxylase -

PS Claim 21; Fig 9; 21pp; German.

CC The present invention relates to a method of influencing the level of  
 CC serotonin, which involves the specific regulation of TPH (tryptophan  
 CC hydroxylase) and/or neurone-specific TPH (nrTPH) activity. The method is  
 CC used for the diagnosis of neuronal diseases and for treating deficient  
 CC primary haemostasis. Also reducing production of serotonin is used to  
 CC treat arteriosclerosis and thrombosis, particularly where associated with  
 CC diabetes mellitus, and to treat excessive (or normal but unwanted) immune  
 CC system responses, e.g. allergy, (auto)immune diseases, risks associated  
 CC with pregnancy, particularly pre-eclampsia, and transplant rejection. The  
 CC present sequence is a neurone-specific TPH protein.

CC Sequence 434 AA;

Query Match 12.0%; Score 226.5; DB 23; Length 434;  
 Best Local Similarity 27.4%; Pred. No. 5.1e-14;  
 Matches 63; Conservative 43; Mismatches 97; Indels 27; Gaps 4;

OY 107 WYRLSSRFLSMKSCYCPREFLDYLAFLGLSDFLDH-----QAVIKFELETHFSY 158  
 DB 164 WGTFOELNKLYPTHACK---EYLKMLPLSKICGIRENITPLEDVSNFKERTGFSIR 220  
 OY 159 PVSGFVAAPHOYLSTLDQRYPIASVMRTLDKDNFSLTPDLIHLLGHVPMILHPSSEFF 218  
 DB 221 PVAGYLSRDFLSGLAFRFVHCTQYVRHSSDPYTPPEPTCHELGHVPLAEPSPAFQS 280  
 OY 219 INMGRLEFTKVIKQVALPSKKORIQTLOSNIATVRCFWFTVESGLIENHGRKAYGAVL 278  
 DB 281 OEIG-----LASLGASEEAVQKLTAT-----CYFFTVFGLCKQDQGLRVFGAGL 324  
 OY 279 ISSPOELGHAFIDNVRVLPLELDQIRLPFNTSPQETLSIRHDELVE 328  
 DB 325 LSSISELKHVLSGHAKVFPDPKITCKQECILTTFQDVYFVESPEDAKE 374

#### RESULT 8

ID AAR50185 standard; Protein; 498 AA.

AC AAR50185;

DT 04-NOV-1994 (first entry)

DE Rat tyrosine hydroxylase.

KW Human; rat; tyrosine hydroxylase; TH; substitution; N-terminal;

KW enhanced enzymatic activity; wild-type; transformation; dopamine;

KW Parkinson's disease; Alzheimer's disease; brain; encapsulation;

KW selectively permeable; polymer capsule; antibody.

OS Rattus rattus.

PN US5300436-A.

PD 05-APR-1994.

PF 13-MAR-1991; 91US-0669446.

PR 13-MAR-1991; 91US-0669446.

PR 26-JAN-1993; 93US-0009075.

PA (UYNV ) UNIV NEW YORK STATE.

PI Filler D, Friedhoff AJ, Goldstein M, Wu J;

DR WPI; 1994-125849/15.

PT Modified DNA encoding variant tyrosine hydroxylase with an

PT N-terminal amino acid substn. and cells contg. it - can be used

PT to treat diseases associated with defective function of the

PT enzyme, e.g. Parkinson's disease or Alzheimer's disease

PS Disclosure; Fig 5; 27pp; English.



CC This sequence represents rat tyrosine hydroxylase (TH) and may be  
 CC used in the production of the variant TH molecules of the invention.  
 CC These variants contain amino acid substitutions in the N-terminal  
 CC portion of the molecule, and in the rat molecule these substitutions  
 CC are pref. at positions Ser8, Ser19, Ser31, Arg38, Glu43 or  
 CC Arg46. Variants containing one or more of these amino acid  
 CC substitutions, have substantially enhanced enzymatic activity  
 CC compared to the wild-type enzyme. The most pref. substitution being  
 CC Ser40 for Tyr or Leu. DNA encoding the TH variants, and cells  
 CC transformed with this DNA may be used for treating diseases associated  
 CC with defective function of TH, or dopamine, eg. Parkinsons disease  
 CC and Alzheimers disease, and effective disorders. The cells can then  
 CC be implanted into the brain or encapsulated in a selectively permeable  
 CC polymer capsule which allows release of the cells products but protects  
 CC them from attack by the hosts antibodies or cells.

XX Sequence 498 AA;

Query Match 11.8%; Score 222; DB 15; Length 498;  
 Best Local Similarity 26.7%; Pred. No. 1.8e-13;  
 Matches 65; Conservative 39; Mismatches 103; Indels 36; Gaps 5;

QY 112 SSFSLMKS-----YCRFPLDYLEAFGLLSDFLDH-----QAVIKFELETHF 155  
 DB 227 AEEATWKEVVTGLGYATHACREHLEGFQLLERYCGYREDSTPQLEDVSRFLKERTGF 286  
 QY 156 SYYPVSGFVAPHQVLSLLODRYFPPIASVMRTLDKDNFSLTPDLIHDLGHVPMLLHPSPS 215  
 DB 287 QLRFPVAGLLSAKDFLASLAFRVFQCTQYIRHASSPMHSPEDPCCHELLGHVPMPLADRTFA 346  
 QY 216 EFFINMGRLFVKIEKQALPSKKQRIQTLOSNIATVRCFWFTVESGLIENHEGRKAYG 275  
 DB 347 QFSQDIG-----LASIGASDEEIEKLSI-----VYVFTVEFGLCKQNGELKAYG 390  
 QY 276 AVLSSSQELGHAFIDNVRVLPLELDQIIRLPFNFTSTPQETLFSIRHF-----DELVELTS 331  
 DB 391 AGLSSYGEHLSLSEPEVRAPDPDTAAVQYQDOTYQPVYFVSEFNDKDKLRNAYAS 450  
 QY 332 KLE 334  
 DB 451 RIQ 453

RESULT 9  
 AAR36741  
 ID AAR36741 standard; protein; 498 AA.

AC AAR36741;  
 DT 20-SEP-1993 (first entry)  
 DE Rat tyrosine hydroxylase.  
 KW Human; rat; tyrosine hydroxylase; substitution; N-terminal; dopamine;  
 KW neurological disorder; antibody; variant.  
 OS Rattus rattus.

Key Location/Qualifiers  
 FT Modified-site 8 /note= "Phosphorylation site"  
 FT Modified-site 19 /note= "Phosphorylation site"  
 FT Modified-site 40 /note= "Phosphorylation site"  
 FT Modified-site 153 /note= "Phosphorylation site"  
 FT Modified-site /note= "Phosphorylation site"

XX US5212082-A.  
 XX 18-MAY-1993.  
 XX 13-MAR-1991; 91US-0669446.

XX 13-MAR-1991; 91US-0669446.  
 XX (UUNY ) UNIV NEW YORK STATE.  
 XX Filler D, Friedhoff AJ, Goldstein M, Wu J;  
 XX WPT; 1993-175456/21.  
 XX Genetically modified tyrosine hydroxylase having increased  
 XX activity - used for treating neurological disorders e.g.  
 XX Parkinson's and Alzheimer's diseases and affective disorders  
 XX Disclosure; Fig 5; 20pp; English.  
 XX The sequences given in AAR36740-41 represent the human and rat  
 XX tyrosine hydroxylase protein respectively. These sequences may be  
 XX used in the construction of a variant tyrosine hydroxylase which  
 XX contains at least one amino acid substitution in the N-terminal 55  
 XX amino acids. The substitution corresponds to an amino acid selected  
 XX from Ser8, Ser31, Arg37, Arg38, Gln39, Ser40, Leu42, Glu43,  
 XX Asp44, Ala45, Arg46 or Lys47. The enzymatic activity of the variant  
 XX proteins is at least 3-fold greater than that for the wild type  
 XX protein. Cells transfected with the DNA encoding these proteins may  
 XX be used for treating neurological disorders associated with a  
 XX deficiency in tyrosine hydroxylase or dopamine. These proteins may  
 XX be used to generate antibodies specific for the variant tyrosine  
 XX hydroxylases to monitor the enzyme during a treatment regimen.

XX Sequence 498 AA;

Query Match 11.7%; Score 221; DB 14; Length 498;  
 Best Local Similarity 27.4%; Pred. No. 2.2e-13;  
 Matches 65; Conservative 37; Mismatches 99; Indels 36; Gaps 5;

QY 118 WKS-----YCRFPLDYLEAFGLLSDFLDH-----QAVIKFELETHFSYYPVS 161  
 DB 233 WREVVTGLGYATHACREHLEGFQLLERYCGYREDSTPQLEDVSRFLKERTGFQLRPVA 292  
 QY 162 GFVAPHQVLSLLODRYFPPIASVMRTLDKDNFSLTPDLIHDLGHVPMLLHPSPSEFFINM 221  
 DB 293 GLLSAKDFLASLAFRVFQCTQYIRHASSPMHSPEDPCCHELLGHVPMPLADRTFAQFSQDI 352  
 QY 222 GRLETKVIEKQALPSKKQRIQTLOSNIATVRCFWFTVESGLIENHEGRKAYGAVLISS 281  
 DB 353 G-----LASIGASDEEIEKLSI-----VYVFTVEFGLCKQNGELKAYGAGLLSS 396  
 QY 282 PQELGHAFIDNVRVLPLELDQIIRLPFNFTSTPQETLFSIRHF-----DELVELTSKLE 334  
 DB 397 YGELLHLSLSEPEVRAPDPDTAAVQYQDOTYQPVYFVSEFNDKDKLRNAYASRIQ 453

RESULT 10  
 AAR50184  
 ID AAR50184 standard; Protein; 497 AA.

AC AAR50184;  
 DT 04-NOV-1994 (first entry)  
 DE Human tyrosine hydroxylase.  
 KW Human; rat; tyrosine hydroxylase; TH; substitution; N-terminal;  
 KW enhanced enzymatic activity; wild-type; transformation; dopamine;  
 KW Parkinsons disease; Alzheimers disease; brain; encapsulation;  
 KW selectively permeable; polymer capsule; antibody.  
 OS Homo sapiens.  
 XX US5300436-A.  
 XX 05-APR-1994.

PF 13-MAR-1991; 91US-0669446.  
 XX 13-MAR-1991; 91US-0669446.  
 PR 26-JAN-1993; 93US-0009075.  
 XX (UWNY ) UNIV NEW YORK STATE.  
 PA Filer D, Friedhoff AJ, Goldstein M, Wu J;  
 PI WPI; 1994-125849/15.  
 DR N-PSDB; AAQ44699.  
 XX Modified DNA encoding variant tyrosine hydroxylase with an  
 PT N-terminal amino acid subseq. and cells contg. it - can be used  
 PT to treat diseases associated with defective function of the  
 PT enzyme, e.g. Parkinson's disease or Alzheimer's disease  
 XX  
 XX Disclosure; Column 21-24; 27pp; English.  
 PS  
 CC This sequence represents human tyrosine hydroxylase (TH) and may be  
 CC used in the production of the variant TH molecules of the invention.  
 CC These variants contain amino acid substitutions in the N-terminal  
 CC portion of the molecule, and in the human molecule these  
 CC substitutions are pref. at positions Ser19, Ser31, Arg37, Arg38,  
 CC Ser40, Leu41, Glu43, Asp44, Ala45, Arg46 or Lys47. Variants  
 CC containing one or more of these amino acid substitutions, have  
 CC substantially enhanced enzymatic activity compared to the wild-type  
 CC enzyme. The most pref. substitution being Ser40 for Tyr or Leu.  
 CC DNA encoding the TH variants, and cells transformed with this DNA  
 CC may be used for treating diseases associated with defective function  
 CC of TH, or dopamine, eg. Parkinsons disease and Alzheimers disease,  
 CC and effective disorders. The cells can then be implanted into the  
 CC brain or encapsulated in a selectively permeable polymer capsule  
 CC which allows release of the cells products but protects them from  
 CC attack by the hosts antibodies or cells.  
 XX  
 XX Sequence 497 AA;  
 SQ  
 Query Match 11.6%; Score 220; DB 15; Length 497;  
 Best Local Similarity 26.7%; Pred. No. 2.8e-13;  
 Matches 65; Conservative 38; Mismatches 104; Indels 36; Gaps 5;  
 QY 112 SSRFSLWKS-----YCPFFLDYEAFGLLSPDLH-----QAVIKFFELTHP 155  
 DB 226 AEEIATWKEVYTTLKGIVATHACGHELEAFALREFSGYREDNIPQLEDVSRLEKRTGP 285  
 QY 156 SYYPVSGFVAAPHQVLSLQDRYFPPIASVMTLKDNPFLPDLIHDLGHVPLLPSPFS 215  
 DB 286 QLRPVAGLLSARDPLASLAFVFOCTQYIRHASSPMHSPEPDCHELLGHVPLMDRTFA 345  
 QY 216 EEFINMGRLLFTKVIKQVQALPSKQRIQTLOSNIILAIIVRCFMTVESGLIENHEGRKAYG 275  
 DB 346 QFSQDIG-----LASLGASDEIEKLTLS-----WFTVEFGLCQNGEVKAYG 389  
 QY 276 AVLISSPOELGHAFFIDNVLEPLLEDOIIRLPFNTSPQETLFSIRNF---DELVELTS 331  
 DB 390 AGLISSYGEILHCLSEEPETIRAFDEPAAVQPYQDOTQSVYVSSFSFSDAKDKLSRYAS 449  
 QY 332 KLE 334  
 DB 450 RIQ 452  
 RESULT 11  
 AAR36740  
 ID AAR36740 standard; protein; 497 AA.  
 XX AAR36740;  
 AC AAR36740;  
 XX 20-SEP-1993 (first entry)  
 DT Human tyrosine hydroxylase.  
 XX  
 DE

KW Human; rat; tyrosine hydroxylase; substitution; N-terminal; dopamine;  
 XX neurological disorder; antibody; variant.  
 XX Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 19  
 FT Modified-site /note= "Phosphorylation site"  
 FT Modified-site 40  
 FT Modified-site /note= "Phosphorylation site"  
 FT Modified-site 152  
 FT Modified-site /note= "Phosphorylation site"  
 FT Modified-site 222  
 FT Modified-site /note= "Phosphorylation site"  
 XX US5212082-A.  
 XX 18-MAY-1993.  
 PD  
 XX 13-MAR-1991; 91US-0669446.  
 PF 13-MAR-1991; 91US-0669446.  
 XX 13-MAR-1991; 91US-0669446.  
 PR (UWNY ) UNIV NEW YORK STATE.  
 PA Filer D, Friedhoff AJ, Goldstein M, Wu J;  
 PI WPI; 1993-175456/21.  
 DR N-PSDB; AAQ41886.  
 XX Genetically modified tyrosine hydroxylase having increased  
 PT activity - used for treating neurological disorders e.g.  
 PT Parkinson's and Alzheimer's diseases and affective disorders  
 XX  
 XX Disclosure; Fig 6; 20pp; English.  
 PS  
 CC The sequences given in AAR36740-41 represent the human and rat  
 CC tyrosine hydroxylase protein respectively. These sequences may be  
 CC used in the construction of a variant tyrosine hydroxylase which  
 CC contains at least one amino acid substitution in the N-terminal 55  
 CC amino acids. The substitution corresponds to an amino acid selected  
 CC from Ser3, Ser31, Arg37, Glu43, Ser40, Leu41, Glu42, Glu43,  
 CC Asp44, Ala45, Arg46 or Lys47. The enzymatic activity of the variant  
 CC proteins is at least 3-fold greater than that for the wild type  
 CC protein. Cells transfected with the DNA encoding these proteins may  
 CC be used for treating neurological disorders associated with a  
 CC deficiency in tyrosine hydroxylase or dopamine. These proteins may  
 CC be used to generate antibodies specific for the variant tyrosine  
 CC hydroxylases to monitor the enzyme during a treatment regimen.  
 XX  
 XX Sequence 497 AA;  
 SQ  
 Query Match 11.6%; Score 219; DB 14; Length 497;  
 Best Local Similarity 26.7%; Pred. No. 3.5e-13;  
 Matches 65; Conservative 38; Mismatches 104; Indels 36; Gaps 5;  
 QY 112 SSRFSLWKS-----YCPFFLDYEAFGLLSPDLH-----QAVIKFFELTHP 155  
 DB 226 AEEIATWKEVYTTLKGIVATHACGHELEAFALREFSGYREDNIPQLEDVSRLEKRTGP 285  
 QY 156 SYYPVSGFVAAPHQVLSLQDRYFPPIASVMTLKDNPFLPDLIHDLGHVPLLPSPFS 215  
 DB 286 QLRPVAGLLSARDPLASLAFVFOCTQYIRHASSPMHSPEPDCHELLGHVPLMDRTFA 345  
 QY 216 EEFINMGRLLFTKVIKQVQALPSKQRIQTLOSNIILAIIVRCFMTVESGLIENHEGRKAYG 275  
 DB 346 QFSQDIG-----LASLGASDEIEKLTLS-----WFTVEFGLCQNGEVKAYG 389  
 QY 276 AVLISSPOELGHAFFIDNVLEPLLEDOIIRLPFNTSPQETLFSIRNF---DELVELTS 331  
 DB 390 AGLISSYGEILHCLSEEPETIRAFDEPAAVQPYQDOTQSVYVSSFSFSDAKDKLSRYAS 449  
 QY 332 KLE 334

[illegible]

Matches 62; Conservative 44; Mismatches 100; Indels 27; Gaps 4;

QY 104 RNLWRLSSRPSLMKSYCPREFLDYLAEGLSLSDLDH-----QAVIKFELETHF 155  
 DB 184 KKTWGVTFKTLKSLYKTHA---CYEYNHIFPLEKCYGFHEDNIPOLLEVDVSOFLQTCGF 240  
 QY 156 SYYPVSGFVAPHQYLSLQDRYFPFASVWRTLDKNSLTPDLIHDLGHVPMLLHPSFS 215  
 DB 241 RLRPVAGLLSRDPLGGLAFRVFHCQYIRHGSKPMWTPPEPICHELLGHVPLFSDRSFA 300  
 QY 216 EPPINMGRLEFTKVIKYOALPSKKORIQTLOSNIIVRCFMFTVESGLIENHGRKAYG 275  
 DB 301 QFSQEIQ-----LASLGAPEYIEKLTAT-----IYWFTVEFGLCCKGDSIRKAYG 344

QY 276 AVLISSPQELGHAFIDNVAVPLELDQIIRLPNTSTPQETLSIRHPELVE 328  
 DB 345 AGLISSFGELOYCLSEKPKLPLELEKTAIONVTVEFPQPLVYVASFNDAKE 397

RESULT 14  
 AAW25788 ID AAW25788 standard; Protein; 452 AA.  
 XX AAW25788;  
 AC AAW25788;  
 XX 27-MAR-1998 (first entry)  
 DT  
 XX Human phenylalanine hydroxylase.  
 DE  
 XX Phenylalanine hydroxylase; human; hybrid protein; cell delivery;  
 KW cell binding ligand; translocation domain; diphtheria toxin B';  
 KM phenylketonuria; therapy.  
 XX  
 XX Homo sapiens.  
 OS  
 PN US5668255-A.  
 XX 16-SEP-1997.  
 PD  
 XX 07-JUN-1984; 84US-0618199.  
 PF  
 XX 27-JUN-1991; 91US-0722484.  
 PR 07-JUN-1984; 84US-0618199.  
 PR 25-APR-1985; 85US-0726808.  
 PR 07-JUN-1985; 85US-0742554.  
 PR 22-DEC-1989; 89US-0456095.  
 PR 14-JUN-1990; 90US-0538276.  
 PR 04-AUG-1993; 93US-0102387.  
 XX  
 XX (SERA-) SERAGEN INC.  
 PA  
 XX Murphy JR;  
 PI  
 XX WPI; 1997-470103/43.  
 DR N-PSDB; AAT91639.  
 XX  
 PT New hybrid molecules for delivery of agents to cells - comprise a  
 PT binding domain of a cell binding ligand and a portion of a  
 PT translocation domain of a protein  
 XX  
 XX Example 5; Fig 13A-C; 30pp; English.

XX This protein comprises for human phenylalanine hydroxylase (PH).  
 CC A DNA molecule (see AAT91638) encoding PH was used to construct a  
 CC PH-diphtheria toxin B' gene that was expressed in E. coli. The  
 CC resulting hybrid protein can be used in the treatment of  
 CC phenylketonuria. The active PH enzyme is targeted to, and  
 CC incorporated into, the broad range of cells which native diphtheria  
 CC normally attacks, achieving the widespread therapy that is needed  
 CC for this inherited disorder. Claimed hybrid proteins comprise a  
 CC translocation domain and a cell binding domain. They can be used  
 CC for the delivery of agents (e.g. therapeutic genes, toxins,  
 CC detectable labels) into cells. The use of a translocation

CC mechanism ensures that the hybrid will be effective in relatively  
 CC low doses, since a high proportion of the substance of interest  
 CC will be taken into the targeted cells. The hybrid molecules can be  
 CC manufactured as a single hybrid recombinant protein, permitting  
 CC reproducibility, consistency, and the precise control of  
 CC composition.

QY 104 RNLWRLSSRPSLMKSYCPREFLDYLAEGLSLSDLDH-----QAVIKFELETHF 155  
 DB 184 KKTWGVTFKTLKSLYKTHA---CYEYNHIFPLEKCYGFHEDNIPOLLEVDVSOFLQTCGF 240  
 QY 156 SYYPVSGFVAPHQYLSLQDRYFPFASVWRTLDKNSLTPDLIHDLGHVPMLLHPSFS 215  
 DB 241 RLRPVAGLLSRDPLGGLAFRVFHCQYIRHGSKPMWTPPEPICHELLGHVPLFSDRSFA 300  
 QY 216 EPPINMGRLEFTKVIKYOALPSKKORIQTLOSNIIVRCFMFTVESGLIENHGRKAYG 275  
 DB 301 QFSQEIQ-----LASLGAPEYIEKLTAT-----IYWFTVEFGLCCKGDSIRKAYG 344

QY 276 AVLISSPQELGHAFIDNVAVPLELDQIIRLPNTSTPQETLSIRHPELVE 328  
 DB 345 AGLISSFGELOYCLSEKPKLPLELEKTAIONVTVEFPQPLVYVASFNDAKE 397

RESULT 15  
 AAY55893 ID AAY55893 standard; Protein; 452 AA.  
 XX AAY55893;  
 AC AAY55893;  
 XX 15-FEB-2000 (first entry)  
 DT  
 XX Human phenylalanine hydroxylase.  
 DE  
 XX Recombinant; hybrid; binding domain; ligand; animal cell; diphtheria;  
 KW translocation domain; botulinum; neurotoxin; ricin; cholera; tetanus;  
 KW shiga-like toxin; pertussis; translocation; cytoplasmic membrane; HIV;  
 KM cytosol; therapy; genetic deficiency disease; enzyme; co-factor; poison;  
 KM adipocyte; cancer; virus; infection; antibody.  
 XX  
 XX Homo sapiens.  
 OS  
 PN US5965406-A.  
 XX 12-OCT-1999.  
 PD  
 XX 07-JUN-1995; 95US-0488246.  
 PF  
 XX 04-AUG-1993; 93US-0102387.  
 PR 07-JUN-1984; 84US-0618199.  
 PR 27-JUN-1991; 91US-0722484.  
 PR 25-APR-1985; 85US-0726808.  
 PR 07-JUN-1985; 85US-0742554.  
 PR 22-DEC-1989; 89US-0456095.  
 PR 14-JUN-1990; 90US-0538276.  
 XX  
 XX (SERA-) SERAGEN INC.  
 PA  
 XX Murphy JR;  
 PI  
 XX WPI; 1999-632431/54.  
 DR N-PSDB; AAZ30664.  
 XX  
 PT Recombinant DNA molecule encoding a three part hybrid protein used in  
 PT the treatment of AIDS and genetic deficiency diseases -  
 PT Example 5; Fig 13; 31pp; English.

XX The invention relates to a recombinant DNA molecule encoding a hybrid  
 CC protein comprising three parts: (a) the first part comprises a portion  
 CC of the binding domain of a cell-binding polypeptide ligand allowing the  
 CC hybrid protein to bind to an animal cell; (b) the second part comprises  
 CC a portion of a translocation domain of a naturally occurring protein  
 CC selected from diphtheria toxin, botulinum neurotoxin, ricin, cholera  
 CC toxin, LT toxin, C3 toxin, shiga-like toxin, pertussis toxin and tetanus  
 CC toxin, which translocate the third part of the across the cytoplasmic  
 CC membrane into the cytosol of the cell; and (c) the third part comprises  
 CC a polypeptide entity to be introduced into the cell, which is non-native  
 CC to the naturally occurring protein of (b). This sequence represents the  
 CC human phenylalanine hydroxylase for use in generating the hybrid of the  
 CC invention. The hybrid molecule enables the use in generating of appropriate  
 CC therapy to affected cells, allowing them to function properly and  
 CC alleviate or cure the disease. The hybrid is especially used in treating  
 CC genetic deficiency diseases, by delivering to affected cells an enzyme  
 CC supplying the missing function, to supplementing cellular levels of a  
 CC particular enzyme or a scarce precursor or cofactor, to directing toxins  
 CC or other poisons to destroy particular cells (such as adipocytes, cancer  
 CC cell, or virus infected cells), to counteracting viral infections, such as  
 CC HIV, by introducing appropriate antibodies to viral proteins. It is also  
 CC involved in the process of getting non-therapeutic substances such as  
 CC detectable labels into cells.

XX Sequence 452 AA;

Query Match 11.3%; Score 213.5; DB 20; Length 452;

Best Local Similarity 26.6%; Pred. No. 1.1e-12;

Matches 62; Conservative 44; Mismatches 100; Indels 27; Gaps 4;

QY 104 RNLWYRLSSRFLSKSYCPRFFLDYLEAFGLSLDFLDH-----QAVIKFPELETHF 155  
 DB 184 KKTGTGTVFKTLKSLYKTHA---CYEYNHIFPLLEKYCGFHEDNIPQLEDVSQFLQTCCTGF 240  
 QY 156 SYYPVSGFVAPHOYLSLLQDRYPIASVNRITLDKNFSLTDPDLHDLGHVPLWLLHPSFS 215  
 DB 241 RLRPVAGLLSRDPLGGLAPRVPHCTQYIRHSGKPMYTPEDICHELLGHVPLFSDRSFA 300  
 QY 216 EFTINMGRFLTQVIEKVOALPSKKQRIQTLQSNLAIIVRCFWFTVESGLIENHEGRKAYG 275  
 DB 301 QFSQEIG-----LASLGADPEVIEKLA-----IYWFTVEGLCKQGSIKAYG 344  
 QY 276 AVLJSSSQELGHAFIDNVRVLPLELDQIIRLPNTSTPQETLFSIRHFDLVE 328  
 DB 345 AGLSSSGELQYCLSEKPKLLPLELEKTAIQNTVTPEQPLYYVAESFNDAKE 397

RESULT 16

AA78593

ID AA78593 standard; Protein; 452 AA.

XX AA78593;

XX 05-MAY-2000 (first entry)

XX Human phenylalanine hydroxylase protein sequence.

XX Phenylalanine hydroxylase; translocation domain; cell destruction;  
 KW cell binding domain; genetic deficiency disease; cell targeting; cancer;  
 KW adipocyte; enzyme delivery; anti-viral; HIV.

OS Homo sapiens.

PN US6022950-A.

PD 08-FEB-2000.

XX 07-JUN-1995; 95US-0479510.

XX 07-JUN-1994; 84US-0618199.

PR 27-JUN-1991; 91US-0722484.

PR 25-APR-1995; 85US-0726808.

PR 07-JUN-1985; 85US-0742554.  
 PR 22-DEC-1989; 89US-0456095.  
 PR 14-JUN-1990; 90US-0538276.  
 PR 04-AUG-1993; 93US-0102387.  
 PA (SERA-) SERAGEN INC.  
 XX Murphy JR;  
 XX WPI; 2000-160390/14.  
 DR N-PSDB; AAZ90020.  
 DR New two-part hybrid protein comprising a translocation domain and a  
 PT cell-binding domain, for treating genetic deficiency diseases, cancer  
 PT and HIV infections.  
 XX Example 5; Fig 13; 32pp; English.

XX This sequence represents the phenylalanine hydroxylase protein sequence.  
 CC The encoded protein can be included in the hybrid protein of the  
 CC invention and used to destroy or modify the cell that the hybrid protein  
 CC is targeted to. The hybrid protein comprises a first part which is a  
 CC portion of the binding domain of a cell-binding ligand, effective to  
 CC cause the hybrid molecule to bind to a cell of an animal. The second  
 CC part comprises a portion of a translocation domain of a naturally  
 CC occurring protein (e.g. the translocation domain of diphtheria toxin) the  
 CC second part translocates the third part across the cytoplasmic membrane  
 CC and into the cytosol of the cell. The third part comprises a chemical  
 CC entity to be introduced into the cell, where each of the first and third  
 CC part is non-native with respect to naturally occurring protein, and the  
 CC covalent bond attaching the second and third part is cleavable. The  
 CC phenylalanine hydroxylase protein can form part of the third portion of  
 CC the hybrid protein. The cell binding domain binds to a specific cell and  
 CC the translocation domain transfers the hybrid molecule across the cell  
 CC membrane into the cytosol. The third part of the protein, linked to the  
 CC translocation domain through a cleavable bond, can then carry out its  
 CC function. The hybrid molecules are useful for treating genetic deficiency  
 CC diseases by delivering to affected cells an enzyme supplying the missing  
 CC function, to supplement cellular levels of a particular enzyme or a  
 CC scarce precursor or cofactor, to direct toxins or other poisons to  
 CC destroy particular cells (such as adipocytes, cancer cells, or  
 CC virus infected cells), and to counteract viral infections such as HIV by  
 CC introducing into appropriate cells antibodies to viral proteins.

XX Sequence 452 AA;

Query Match 11.3%; Score 213.5; DB 21; Length 452;

Best Local Similarity 26.6%; Pred. No. 1.1e-12;

Matches 62; Conservative 44; Mismatches 100; Indels 27; Gaps 4;

QY 104 RNLWYRLSSRFLSKSYCPRFFLDYLEAFGLSLDFLDH-----QAVIKFPELETHF 155  
 DB 184 KKTGTGTVFKTLKSLYKTHA---CYEYNHIFPLLEKYCGFHEDNIPQLEDVSQFLQTCCTGF 240  
 QY 156 SYYPVSGFVAPHOYLSLLQDRYPIASVNRITLDKNFSLTDPDLHDLGHVPLWLLHPSFS 215  
 DB 241 RLRPVAGLLSRDPLGGLAPRVPHCTQYIRHSGKPMYTPEDICHELLGHVPLFSDRSFA 300  
 QY 216 EFTINMGRFLTQVIEKVOALPSKKQRIQTLQSNLAIIVRCFWFTVESGLIENHEGRKAYG 275  
 DB 301 QFSQEIG-----LASLGADPEVIEKLA-----IYWFTVEGLCKQGSIKAYG 344  
 QY 276 AVLJSSSQELGHAFIDNVRVLPLELDQIIRLPNTSTPQETLFSIRHFDLVE 328  
 DB 345 AGLSSSGELQYCLSEKPKLLPLELEKTAIQNTVTPEQPLYYVAESFNDAKE 397

RESULT 17

AAB13327

ID AAB13327 standard; Protein; 532 AA.

XX AAB13327;

XX

DT 12-JAN-2001 (first entry)  
 XX  
 DE Caenorhabditis elegans Cod-5.  
 XX  
 KM Caenorhabditis elegans; cod-5; daf-18; insulin signalling pathway;  
 KM daf-2; age-1; insulin receptor; PI 3-kinase; PKB kinase;  
 KM PTEN lipid phosphatase; antidiabetic; anorectic; obesity; diabetes.  
 XX  
 OS Caenorhabditis elegans.  
 XX  
 PN WO200033068-A1.  
 XX  
 PD 08-JUN-2000.  
 XX  
 PF 02-DEC-1999; 99WO-US28529.  
 XX  
 PR 03-DEC-1998; 98US-0205658.  
 XX  
 PA (GENO ) GEN HOSPITAL CORP.  
 XX  
 PI Ruvkun G, OGS S;  
 XX  
 DR WPI: 2000-423022/36.  
 DR N-PSDB; AAA91624.  
 XX  
 PT Diagnosing and treating obesity and impaired glucose tolerance using  
 PT modulators of daf-18 expression and/or activity -  
 XX  
 PS Disclosure; Fig 42; 402pp; English.  
 XX  
 CC The present sequence is Cod-5 from Caenorhabditis elegans. Cod-5  
 CC is the aromatic amino acid hydroxylase that synthesises serotonin from  
 CC precursor L tryptophan. The cod-5 gene was knocked out to produce mutants  
 CC in metabolic control. A number of C. elegans proteins that have defects  
 CC in mammalian homologues acting in the insulin signalling pathway were also  
 CC identified. The C. elegans age-1 gene encodes a homologue of the  
 CC mammalian PI 3-kinase whilst daf-2 encodes a homologue of the mammalian  
 CC insulin receptor. The C. elegans AKT kinase and PKB kinase act downstream  
 CC of daf-2 and age-1, just as their mammalian homologues act downstream of  
 CC insulin signalling. The C. elegans PTEN lipid phosphatase homologue,  
 CC DAF-18, has been found to act upstream of AKT in the pathway. This  
 CC discovery has enabled mammalian PTEN action to be mapped to the insulin  
 CC signalling pathway. Conserved DAF motifs can be used to design probes to  
 CC identify mammalian DAF homologues and thus to identify individuals with a  
 CC predisposition towards the development of glucose intolerance conditions,  
 CC such as obesity and diabetes.  
 CC  
 XX  
 SQ Sequence 532 AA;  
 Query Match 10.9%; Score 206.5; DB 21; Length 532;  
 Best Local Similarity 25.2%; Pred. No. 7.1e-12;  
 Matches 70; Conservative 47; Mismatches 116; Indels 45; Gaps 8;  
 QY 104 RNLW---YRLSSRPSLWKSVCPRFFLDYLEAFGLSDFLDH-----QAVIKFPELETHF 155  
 DB 251 RKTWGIYRKLR---ELHKHACKQFLDNFELLERHCGYSENNIPQLEDICFLKAKGTF 307  
 QY 156 SYYPVSGFVAPHOYLSLDQDRYPIASVMTLDKNFSLPDLIHDLGHVPMILHPSFS 215  
 DB 308 RVRPVAGVYLSARDPLAGLAVRVFCTQYVRHADPFYTPEDVTVELHGMALFADPDFA 367  
 QY 216 EPIINNGRLFTVKIEKVALPSKKORIQTQSNLIAIVRCFWFTVSSGI-----265  
 DB 368 QFSQELG-----LASTLASEDEDKLATL-----YFESIIFGLSSDDADSPVK 411  
 QY 266 ---ENHEGKAGAVLISPOELGHAFIDNVRVPLEDITLIPNTSTPQETLPSIRH 322  
 DB 412 ENGSNHERKVTAGLSAGLQAVGSGKTTIRPPRIVVQOECLITTFQSAFYTRN 471  
 QY 323 FDELVELTSGLEMLDQGLLESIPLYNOEKYLSGFEVL 360  
 DB 472 FEBAQO---KLRFMTNNMKRPPIVRYN--PYTESVEVL 504

RESULT 18  
 AAU30703  
 ID AAU30703 standard; Protein; 491 AA.  
 XX  
 AC AAU30703;  
 XX  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Novel human secreted protein #1194.  
 XX  
 KM Human; vaccination; gene therapy; nutritional supplement;  
 KM stem cell proliferation; haematopoiesis; nerve tissue regeneration;  
 KM immune suppression; immune stimulation; anti-inflammatory; leukaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200179449-A2.  
 XX  
 PD 25-OCT-2001.  
 XX  
 PF 16-Apr-2001; 2001WO-US08656.  
 XX  
 PR 18-Apr-2000; 2000US-0552929.  
 PR 26-Jan-2001; 2001US-0770160.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT;  
 XX  
 DR WPI: 2001-611725/70.  
 XX  
 PT Nucleic acids encoding a range of human polypeptides, useful in genetic  
 PT vaccination, testing and therapy -  
 XX  
 PS Claim 20; Page 332; 765pp; English.  
 XX  
 CC The invention relates to novel human secreted polypeptides. The  
 CC polypeptides and antibodies to the polypeptides are useful for  
 CC determining the presence of or predisposition to a disease associated  
 CC with altered levels of polypeptide. The polypeptides are also useful for  
 CC identifying agents (agonists and antagonists) that bind to them. Cells  
 CC expressing the proteins are useful for identifying a therapeutic agent  
 CC for use in treatment of a pathology related to aberrant expression or  
 CC physiological interactions of the polypeptide. Vectors comprising  
 CC the nucleic acids encoding the polypeptides and cells genetically  
 CC engineered to express them are also useful for producing the proteins.  
 CC The proteins are useful in genetic vaccination, testing and  
 CC therapy, and can be used as nutritional supplements. They may be used to  
 CC increase stem cell proliferation; to regulate haematopoiesis; and in  
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;  
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and  
 CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid  
 CC sequences of novel human secreted proteins of the invention.  
 CC  
 XX  
 SQ Sequence 491 AA;  
 Query Match 10.6%; Score 199.5; DB 22; Length 491;  
 Best Local Similarity 28.4%; Pred. No. 3.2e-11;  
 Matches 58; Conservative 36; Mismatches 83; Indels 27; Gaps 5;  
 QY 104 RNLWYRLSSRPSLWKSVCPRFFLDYLEAFGLSDFLDH-----QAVIKFPELETHF 155  
 DB 218 KKTWGTVPKTLKSLVYTHA---CYEYNHIFPLEKYCASHEDNIPQLEDVSGFLQCTGF 274  
 QY 156 SYYPVSGFVAPHOYLSLDQDRYPIASVMTLDKNFSLPDLIHDLGHVPMILHPSFS 215  
 DB 275 RLRFVAGLSRDFGLDAFRVHCQYIRHSGSKMTYPERDICHHLGHVPLPSRSH 334  
 QY 216 EPIINNGRLFTVKIEKVALPSKKORIQTQSNLIAIVRCFWFTVSSGLIENHEGKAYG 275  
 DB 335 QFSQELG-----LASTLAPDSIEKL-----APIYFTVERGLCKQSGDITKAYG 378

QY 276 AVLISSPQELGHAFIDNVRVLPLE 299  
 Db 379 AGLSSRFQCYCLSEKPLLPLE 402

RESULT 19  
 ID ABB61277 standard; Protein; 452 AA.

AC ABB61277;  
 DT 26-MAR-2002 (first entry)  
 DE Drosophila melanogaster polypeptide SEQ ID NO 10623.  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX Drosophila melanogaster.  
 OS WO200171042-A2.  
 PN 27-SEP-2001.  
 XX 23-MAR-2001; 2001WO-US09231.  
 XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 PA (PEKE ) PE CORP NY.  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 DR WPI, 2001-656860/75.  
 DR N-PSDB; ABL05380.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX Disclosure; SEQ ID NO 10623; 21pp + Sequence Listing; English.  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 452 AA;  
 Query Match 10.3%; Score 194; DB 22; Length 452;  
 Best Local Similarity 27.1%; Pred. No. 1e-10;  
 Matches 73; Conservative 39; Mismatches 111; Indels 46; Gaps 10;

QY 107 WYRLSSRFSIMKSCYCRPFLDYLEAFGLSD---FLDH-----QAVIKFELETHFSY 158  
 Db 186 WGIIPRNUTKLYKTHACR---EYHVPFLVNCGRFREDNIPQLEDVSNFRDCTGTLR 242  
 QY 159 PVSGVAPHQVLSLLQDRYPFIASVMTLDKONFSLTDPDLHLGHVPMLLHPSSEFF 218  
 Db 243 PVAGLSRDLFLAGLAFVPHSTQYIRHPSKPMYTPEDVCHLMGHVPLFADPAFAQFS 302  
 QY 219 INMGRLFTKVIKQVALPSKKORIOTLOSNIATVRCFWFTESGLIENHEGRKAYGAVL 278  
 Db 303 QBIG-----LASLGADDDYIEKLST-----IFWFTVEYGVCRQEGELKAYGAGL 346

QY 279 ISSPQELGHAFIDNVRVLPLELDQIIRLPNTSTPQETLFSIRHFDLVELTSLKLEWMLD 338  
 Db 347 LSSYGELEYCLTDKPOLKDFE-----PEVTGV---TKYPITQFQPLYTVADSPFETAKE 396  
 QY 339 QGL--LESIP-----LYNQEKYLSGFEVL 360  
 Db 397 KTIKFANSIPRPFGVRYN--AVTQSVLEV 423

RESULT 20  
 ID ABB71186 standard; Protein; 555 AA.  
 AC ABB71186;  
 DT 26-MAR-2002 (first entry)  
 DE Drosophila melanogaster polypeptide SEQ ID NO 40350.  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX Drosophila melanogaster.  
 OS WO200171042-A2.  
 PN 27-SEP-2001.  
 XX 23-MAR-2001; 2001WO-US09231.  
 XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 PA (PEKE ) PE CORP NY.  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 DR WPI, 2001-656860/75.  
 DR N-PSDB; ABL15289.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX Disclosure; SEQ ID NO 40350; 21pp + Sequence Listing; English.  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 555 AA;  
 Query Match 9.4%; Score 178.5; DB 22; Length 555;  
 Best Local Similarity 32.0%; Pred. No. 5e-09;  
 Matches 47; Conservative 28; Mismatches 55; Indels 17; Gaps 3;

QY 143 QAVIKFELETHFSYYPVSGVAPHQVLSLLQDRYPFIASVMTLDKONFSLTDPDLIDL 202  
 Db 284 QDVSYLKKRKGFLGRVAGVLSRDFLSGLAFRVHCTQYIRHSSDPFTYPPDCCHL 343  
 QY 203 LGHVPWLLHPSSEFFINMGRLFTKVIKQVALPSKKORIOTLOSNIATVRCFWFTVES 262  
 Db 344 LGHMPLLANSFAQFSQIEG-----LASLGASDADIEKLATL-----YFFTVEF 387  
 QY 263 GLIENHEGR-KAYGAVLISSPQELGHA 288

Db 388 GLCKQADSTFKVYGAGLLSVAFLOHA 414

## RESULT 21

AAU30702 standard; Protein; 200 AA.

AAU30702;

18-DEC-2001 (first entry)

Novel human secreted protein #1193.

Human; vaccination; gene therapy; nutritional supplement;

stem cell proliferation; haematopoiesis; nerve tissue regeneration;

immune suppression; immune stimulation; anti-inflammatory; leukaemia.

Homo sapiens.

WO200179449-A2.

25-OCT-2001.

16-APR-2001; 2001WO-US08656.

18-APR-2000; 2000US-0552929.

26-JAN-2001; 2001US-0770160.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT;

WPI; 2001-611725/70.

Nucleic acids encoding a range of human polypeptides, useful in genetic

vaccination, testing and therapy -

Claim 20; Page 332; 765pp; English.

The invention relates to novel human secreted polypeptides. The polypeptides and antibodies to the polypeptides are useful for determining the presence of or predisposition to a disease associated with altered levels of polypeptide. The polypeptides are also useful for identifying agents (agonists and antagonists) that bind to them. Cells expressing the proteins are useful for identifying a therapeutic agent for use in treatment of a pathology related to aberrant expression or physiological interactions of the polypeptide. Vectors comprising the nucleic acids encoding the polypeptides and cells genetically engineered to express them are also useful for producing the proteins. The proteins are useful in genetic vaccination, testing and therapy, and can be used as nutritional supplements. They may be used to increase stem cell proliferation, to regulate haematopoiesis, and in bone, cartilage, tendon and/or nerve tissue growth or regeneration; CC immune suppression and/or stimulation; as anti-inflammatory agents; and CC in treatment of leukaemia. AAU29510-AAU3304 represent the amino acid sequences of novel human secreted proteins of the invention.

Sequence 200 AA;

Query Match 8.1%; Score 153; DB 22; Length 200; Best Local Similarity 34.9%; Pred. No. 4, 3e-07; Matches 37; Conservative 21; Mismatches 32; Indels 16; Gaps 2;

197 DLHDLGHVPMILHPSSEFFIMNGRLFTVIEKVALPSKQRQTQSLNIVICF 256

15 DICHETLGHVPLPSRPSAQSQETG-----LASLQAPDVEIKLAT-----IT 58

257 WFTVESGLIENHGRKAYGAVLSSPOELGHAFIDNVRLPLEDQ 302

59 WFTVERGLCKGDSIKAYGAGLSSFPELOYCLSEKPLPLELEK 104

## RESULT 22

AAE11092 standard; peptide; 64 AA.

AAE11092;

18-DEC-2001 (first entry)

phnAB fusion peptide #3.

phenylalanine hydroxylase; PAH; phnA; 4a-carbinolamine dehydratase; phnB;

aromatic aminotransferase; phnC; milk protein; animal protein; casein;

proteinaceous food product; globulin; whey protein; phenylketonuria;

PKU; inherited metabolic disorder; impaired brain function; nootropic;

cell therapy; phnAB fusion peptide.

Undetected.

WO200168822-A2.

20-SEP-2001.

14-MAR-2001; 2001WO-DK00172.

14-MAR-2000; 2000US-0525116.

(NILA-) NILAB APS.

Johnsen M, Ravn P, Madsen SM, Vrang A, Israelsen H, Bredmose L,

Arnau J, Jensen SH, Gjetting T, Nielsen E;

WPI; 2001-590055/66.

Novel recombinant cells comprising a nucleic acid encoding a gene product having phenylalanine hydroxylase activity, that is derived from a prokaryotic organism, is useful for treating phenylketonuria in mammals -

Disclosure; Page 19; 91pp; English.

The patent discloses novel cells comprising a nucleic acid encoding a gene product having phenylalanine hydroxylase (PAH) activity such as phenylalanine hydroxylase (phnA), 4a-carbinolamine dehydratase (phnB) and aromatic aminotransferase (phnC), which are derived from a prokaryotic organism. The patent also relates to fusion proteins comprising a protein emanating and/or stabilising the PAH activity in addition to PAH activity. The cells are also useful for producing PAH. The sequences of the invention are also useful as a milk protein derived from casein, globulin or a whey protein) having reduced content of phenylalanine. The method involves contacting the food product starting material with the cells or fusion proteins such that at least part of the phenylalanine content of the starting material is converted into compounds that do not cause phenylketonuria (PKU) by the enzymatically active product. PAH enzyme is useful for manufacturing a medicament for treating PKU, which is an inherited metabolic disorder resulting in an accumulation in the body of L-phenylalanine and metabolites that can cause impaired brain function. The present sequence is a peptide sequence of phnAB fusion protein.

Sequence 64 AA;

Query Match 5.8%; Score 109.5; DB 22; Length 64; Best Local Similarity 35.6%; Pred. No. 0.002; Matches 21; Conservative 15; Mismatches 22; Indels 1; Gaps 1;

166 PHQ-YSLQDRYPIASVWRTLDKNSLPDLHDLGHVPMILHPSSEFFIMNGR 223

6 PFGTFELASQGFVATRTPEEDLDYLOBPDIHFHFCPLTPMPAEFTHTYCK 64

## RESULT 23



ABP06168  
ID ABP06168 standard; Protein; 104 AA.  
XX AC  
XX ABP06168;  
XX DT 24-JUN-2002 (first entry)  
XX DE Human ORFX protein sequence SEQ ID NO:12318.  
XX DE Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
KW hypertension; hypothyroidism; cholesterol ester storage disease;  
KW immune deficiency; immune disorder; infectious disease;  
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
KW myasthenia gravis.  
XX OS  
XX Homo sapiens.  
XX WO200192523-A2.  
XX PN  
XX 06-DEC-2001.  
XX PD  
XX 29-MAY-2001; 2001WO-US10836.  
XX PF  
XX 30-MAY-2000; 2000US-206132P.  
XX PR  
XX 29-AUG-2000; 2000US-228716P.  
XX PR  
XX (CUFA-) CURAGEN CORP.  
XX PA  
XX Shinkets RA, Leach MD;  
XX PI  
XX WPI: 2002-106308/14.  
XX DR  
XX N-PSDB; ABN21920.  
XX DR  
XX Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and autoimmune disorders -  
XX  
XX Disclosure; SEQ ID 12318; 1037pp; English.  
XX  
XX The present invention describes substantially purified human proteins  
CC (referred to as open reading frame, ORFX, where X is 1-1491 (see Table 1  
CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
CC treating or preventing a pathology associated with an ORFX-associated  
CC disorder in humans, and in the manufacture of a medicament for treating a  
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
CC sequences can be used in gene therapy. ORFX sequences can be used in the  
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
CC storage disease, various immune deficiencies and disorders, infectious  
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
CC arthritis, autoimmune thyroiditis, myasthenia gravis, Graft-versus-host  
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
CC bone degenerative disorders, or periodontal disease, and for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues and conditions resulting from  
CC systemic cytokine damage.  
XX N.B. the sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 104 AA;  
XX SQ

Query Match 5.8%; Score 109.5; DB 23; Length 104;  
Best Local Similarity 28.6%; Pred. No. 0.0041;  
Matches 26; Conservative 20; Mismatches 44; Indels 1; Gaps 1;

QY 125 FF-----LDYLEAFGLSDPLDHOAVIKPFEELETHFSYYPVSGFVAPHOYLSDORRFP 179  
Db 943 SHHSDDVVOGQMKALGVLTVIARQ---KGFS-----SRLLPDEFLRRYQFLAFDDE--P 992  
QY 180 TSWRTLDKDNFSLTPDLIHDL-----LGHVPMILHPSFSFPIINMGRLPTKYIEKYQ 233  
Db 993 V-----EMTKDNCRLL--LFLRLKMEGMALGKTKVFLRYNDEF--LAKLLELOVKRY- 1040  
QY 234 ALPSKKORIQTLSNLIIVRCFWFTYESGLI-----ENHGRKAYGAVLISPDEL 285  
Db 1041 -----IKVQSMRRLALRRK---VKGKVFRLGKGGPEHHD-----VASKIQKA 1082  
QY 286 GHAFIDNVRVLEPLDQIIRL-----PFNTSTPQETLFSI-----RHFDVLELT 330  
Db 1083 FFGFDRVRLPPLVNEKSGQLENTADFIREFAKWKEKSIQGVLLHYRAARFODFVNL 1142  
QY 331 SKLEWMLDQGLLESIPLYNQ 350  
Db 1143 -----QGVHIYNQ 1150

## RESULT 25

ABR35685  
ID ABR35685 standard; Protein; 1782 AA.

XX ABR35685;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 23847.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

PA Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL09788.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX Disclosure; SEQ ID NO 23847; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA

CC sequences (AB101840-AB116175) and the encoded proteins

CC (AB857737-AB872072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

Query Match 5.5%, Score 103.5, DB 22; Length 1782;

Best Local Similarity 25.0%; Pred. No. 0.99;  
Matches 73; Conservative 32; Mismatches 96; Indels 91; Gaps 15;  
QY 67 ILLEP-----FKNLLFYHLISLTKNREGCSNDMAV-----VSNPFNRRLWYR 109  
Db 574 ILLEPQAIYNNHYRNLFFSPKELNFSRASSANINVRQWAGERPDAVAVAI- 632  
QY 110 LLSRFLMKSCYGRFFLDYLEAGLS-----DFLDHOAVIKFELETH-----PSYRP 159  
Db 633 -----GKSCPKF--STAFYAVYNNKCSFIDEIKIALPASIKHHLLFTIYH 681  
QY 160 VSGFVAP-----HOYLSLLQRIFFPIS-----VMTLDKDNFSLTPDLIHDLL 203  
Db 682 VSCCKKFPDILQSPVETPIGYTWLPLEDGLKFKGEPNLPMVWSPPENVSFTIPNVH--L 739  
QY 204 GHVPMILHPSFSFPIINMGRLFTKYIEKYQALPSKKORIQTLSNLI-IVRCFWFTYES 262  
Db 740 FGIKML-----DNHRAVFSINVEAVTA-----IHTLDSFLDRFPLICEYIDTN 783  
QY 263 GLIENHGRKAYGAVLISPDELGHAFID-----NVRVLEPLDQIIRL 306  
Db 784 --IPSHIGEN-----NIETELKKCLDIEYANRBPVRLPLVDKLEL 826

## RESULT 26

ABR35686  
ID ABR35686 standard; Protein; 563 AA.

XX ABR35686;

DT 24-JUL-2002 (first entry)

DE Fungal ZBC protein sequence #112.

XX Secondary metabolite; fungus; ZBC gene; zinc binuclear cluster protein;

KW antibacterial; beta-lactam; anti-hypercholesterolaemic; lovastatin;

KW mevastatin; immunosuppressant; cyclosporin A; ergot alkaloid; ergotamine;

KW angiotensin inhibitor; ovalicin; glucan synthase inhibitor; gliotoxin;

KW fungal toxin; cell surface receptor; plant growth regulator; pigment;

XX insecticide; antineoplastic.

XX Unidentified.

OS WO200224865-A2.

PN 28-MAR-2002.

PD 19-SEP-2001; 2001WO-US29288.

PR 19-SEP-2000; 2000US-233564P.

XX (MICR-) MICROBIA INC.

PA Holtzman D, Madden K, Maxon M, Sherman A;

XX WPI; 2002-352005/38.

DR N-PSDB; ABR79875.

XX New method for improving the production of a secondary metabolite e.g.

PT antineoplastic agent, ergot alkaloid from a fungus involves modulation

PT of the expression of at least one zinc binuclear cluster protein gene

XX Disclosure; SEQ ID 236; 49pp + sequence listing; English.

XX The invention relates to improving the production of a secondary

CC metabolite by a fungus. This involves modulating the expression of at

CC least one ZBC (zinc binuclear cluster protein) gene in a manner to

CC improve the yield of the secondary metabolite. Methods of the invention

CC may be used for improving the production of the secondary metabolite e.g.

CC antineoplastic (such as beta-lactam), an anti-hypercholesterolaemic (such

as lovastatin or mevastatin), an immunosuppressant (such as cyclosporin A),

an ergot alkaloid (such as ergotamine), an angiotensin inhibitor (such

as ovalicin), a glucan synthase inhibitor, gliotoxin family of compounds a fungal toxin, a modulator of cell surface receptor signalling; a plant growth regulator, a pigment, an insecticide, or an antineoplastic compound. The method results in a decrease in fermentor run-time, a decrease in the size of the fermentor required for the production of equivalent amounts of the secondary metabolite, or a decrease in the biomass required for the production, which translates into decreased waste that must be handled in downstream processing. The sequences given in records ABP3575-ABP3722 represent ZEC proteins.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained directly from WIFO at [ftp.wipo.int/pub/published/pct\\_sequences](http://wipo.int/pub/published/pct_sequences).

Sequence	563 AA;
SQ	

Query Match 5.4%; Score 102; DB 23; Length 563;  
Best Local Similarity 21.1%; Pred. No. 0.27;  
Matches 80; Conservative 56; Mismatches 142; Indels 102; Gaps 16;

```

QY      7  TLDPKYLKIALKRLQSLFQNGQSQRAYSTPTYSYRILQENKEKQALARRHCIS 66
      |||
      |||
      |||
Db     185  TLPIIAATQISDLDPVILNFYNSAGI-----PLESSRLINKLINSISQEQYKHLCLUP 239
      |||
      |||
      |||
QY     67  ILEPFFNLLFVHLLLSKQNRGCGCTDM-----AVSTPFFNRNLWYRLLSRFSLWKS 120
      |||
      |||
      |||
Db    240  DXEIIQMLLRAYATKFRTRIRGVNIDLCSRHVSTLVTPLF--QVTEKIQKNTSDLWFA 297
      |||
      |||
      |||
QY    121  YCPREFFLYEAFGLSLDFLDHOAVIKFEFELETHFSYYPVSGFVAPHOYLSLLQDRYFPI 180
      |||
      |||
      |||
Db    298  LC-----EIDGL-----ECVLKY-----RPEFQHDYGRKLP-----325
      |||
      |||
      |||
QY    181  ASVMRTLDKNFSLTDPDLIDLHLLGHVPWLLHPSESEFFINMGRLFTKVIKQALPSKKQ 240
      |||
      |||
      |||
Db    326  ---LRCPFNDDISYN---PHLLLGR---LDCGVSITKSVHSILTSKFDLKLSEYSQLS 376
      |||
      |||
      |||
QY    241  RI-----QTLQSNLIAIVRCFW-----FTVESGLIENHEGRKAYGAV 277
      |||
      |||
      |||
Db    377  LILVDIEAKFYDPSNEDIQFRIYIFLKMVFMTARVNLVYQCFTILDSGILEDBE-----T 429
      |||
      |||
      |||
QY    278  LISSPOELGHAFIDNVRVLFLBLDQIIRLPTNTSTQETLFSIRHFDLVELTSLKLEWML 337
      |||
      |||
      |||
Db    430  IIGN---LGESCIQCVRLL---ISOITIL-----EKRGLLVALLLEIIHALMLAAFCR 476
      |||
      |||
      |||
QY    338  DOGL-----LESIPLYNQEK 352
      |||
      |||
      |||
Db    477  DKGFVEPSDLGDTILYQVR 496
      |||
      |||
      |||

```

RESULT 27  
AAB94081  
ID AAB94081 standard; Protein; 456 AA.

XX  
AC AAB94081;

XX	DT	26-JUN-2001	(first entry)

DE Human protein sequence SEO IP NO:14279.

XX  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX  
OS Homo sapiens.XX  
PN EP1074617-A2-XX  
PD 07-FEB-2001XX  
PF  
28-III-2000: 2000EB-0116126

XX  
DB 20-III-1000. 0010 024002C

PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-011077C

PR 02-MAY-2000; 2000JP-0183767.

15501470-300007 10007 4000 2000 2000

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -

Claim 8; SEQ ID 14279; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH35893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 456 AA;

Query Match 5.3%; Score 101; DB 22; Length 456;  
Best Local Similarity 23.3%; Pred. No. 0.25;  
Matches 81; Conservative 41; Mismatches 98; Indels 128; Gaps 20;

QY 63 KCISILEFF-KNLLFVHLLSLSKNQREGCSTDMAVSTPFFNRNLWYRLLSSRFSLWKS 121

Db 70 KCLELVEHFGPNELRKVLVMLAAQSRR-----SVPLL-RAISYHLVQKPFSLTKD- 118

QY	122	CPRFFLDYLEAFGLLS-----DFLDHO-----AVIKFFEELETH	154
----	-----	----------------------------------------------	-----

```

Db      119  ---VLLDVAYAYGKLSFHOTQVRSORLATDLLSLMPSLTSGEVAHCAKSFALLKWLSL--- 172

```

Ov 155 FSYYPVSGEVAPHOYLSLLODRYFP-IASVMRTLDKDNFSLTPD-----LIHDLIG-- 204

db 173 ----PLEGAFAOH-VLNRAODITLPHI.CSV.I.AFARI.NFH--PDQEDOFFSI.VHEKI.GSE 225

Ov 205 -----HVPW-----I.I.HPSESEFEINMCPI.ETKVIEKVOAI.PSKK 239

226 I BCI EPRI QVDTI WMAI CWT OQAPREAEI CAWU HDBB  
: : : :  
: |

HTOEY CCYS 269  
: | | |

240 QBTOT CONT T A TUDCEWETTESCY T ENUEECBVA VCAW TCCDCEI CHAET DANEDVY DYE 200

[illegible][illegible][illegible][illegible]

**RESULT 28**

ABB59689  
ID ABB59689 standard; Protein; 2042 AA.



```

Db 233 KVEGSMQKLEN--VLNRASNTADTLFOEVLGRKDK--ADSTRNALNVLRFK-FLFNL 286
Qy 79 LLSLSKQREGCSTDMVAVSTPFENRWYLLLSRF-----SLWKSYPFRPFDLYLEAF 133
Db 287 PLNIKRNIQKG--DYDVVIND-----YEKAKSLFGKTEVQVFKY-----YAEVE 329
Qy 134 GLLSDFLDHQAIVKFELETHESYYPVSGFV-----APHQY-LSLIQ 174
Db 330 AGIEDL--RELLKKL-LETPSLTHDQKRYIRYLSDLHAPCDPAWQCIGAHQWIKLQ 386
Qy 175 D-----RYFPIASVMRTLDKDNESLTPDLIHDLLGH----- 206
Db 387 DCKEGHMKSLKGNPGHSPM-----LDLDN-----DARPSVLGHLSQTASLKRGSSTQS 435
Qy 207 -----PWLH-----PSFSEFFINM--GRFTKVIK-----VQALPSK 238
Db 436 GRDTRWRYKTHRAVFEKLTKLIVLSQLPNFKWISYVNGSLFSETAKSGQIERSKNV 495
Qy 239 KOR-----IQTLOSNLIAVR--CFWFTVESGLIENHEGRKAYGAVLIS---SPQELG 286
Db 496 RORQNDFFKMIQEVHSLVKLIRGALLPFLREG-----DGRQ-YGGEVQVQALSGQWLA 549
Qy 287 HAFIDNVRV-----LFLDLQIIRLPNTSTPQETLFSIR-----HFDLVL 327
Db 550 HV-IQTIRLVYESLTALEIPNDMLQII-----QDLILDLRIHCIMVTLQHTAEIK 599
Qy 328 ELTSKLEWMLDQGLLESIPLYNQEKVLSGFVL 360
Db 600 RLAEKEDWIDNEGLTSLPCQFEQSIHVSLQSL 632

```

## RESULT 30

```

AAW19604
ID AAW19604 standard; Protein; 1024 AA.
AC AAW19604;
XX
DT 21-AUG-1997 (first entry)
DE Mycoplasma genitalium 116 kDa protein MG075 useful in vaccine.
KW Mycoplasma; immunogen; vaccine; diagnosis; pneumonia; inflammation.
OS Mycoplasma genitalium.
PN WO9721727-A1.
PD 19-JUN-1997.
PF 13-DEC-1996; 96WO-AU00803.
PR 13-DEC-1995; 95AU-0007127.
PA (UYME ) UNIV MELBOURNE.
PI Browning GF, Duffy MF, Walker ID, Whithear KG;
XX WPI; 1997-332722/30.
DR
XX New immunogenic polypeptide(s) from Mycoplasma species - useful in
PT vaccines and for diagnosis of Mycoplasma infection
XX
PS Claim 19; Page 85-89; 110pp; English.
XX
CC Isolated or recombinant immunogenic polypeptides from Mycoplasma
CC genitalium have mol.wt. of 16 kDa (AAW19603) (MG074) and 116 kDa
CC (AAW19604) (MG075). They are homologues of 16 and 116 kDa proteins
CC (see also AAW19601-02) obtd. from Mycoplasma pneumoniae. A genomic
CC DNA sequence of M. genitalium contains contiguous open reading
CC frames that code for the 2 polypeptides. Mycoplasma 16 or 116 kDa
CC proteins, or immunogenic fragments that include a T or B cell
CC epitope, can be used in vaccines for prevention and treatment of

```

```

CC Mycoplasma infections, partic. in humans. They can also be used
CC diagnostically to detect Mycoplasma, or to raise antibodies useful
CC in immunoassays.
SQ Sequence 1024 AA;
Query Match 5.2%; Score 98.5; DB 18; Length 1024;
Best Local Similarity 19.9%; Pred. No. 1.4;
Matches 68; Conservative 41; Mismatches 134; Indels 99; Gaps 11;
Qy 29 QNSOSIORAVSTYSYRRILOKENKEK---QALARHKCISILEFPKNLLFVHLLSLSKN 85
Db 564 QQTDSLKNLFSV---IGDILSETNVNKTITHAVNNELLSLVETASTLKIKHL----- 613
Qy 86 QREGCSTDMAVSTPFENRWYLLLSRFSIMKSYCFRFLDYLEAFGLLSDFLDHQAV 145
Db 614 -----NVQYKVLVDKFKLKNSTFK-----ELLNFFPDTKDITPT 647
Qy 146 IK--FFELETH-----FSYYPVSGFVAPHQVYLSLQDRYFPPIASVMRTLDKNF 192
Db 648 IKKVLFESENYKTLRKKEYENEGPGYHWAKEFIVPGTFNSAENTFYSAI-----DKT 698
Qy 193 SLTPDLIHDLLGHVPWLLHPSFSEFFINNGRLFT-----KVIEKVALPSKKQRIQT 244
Db 699 KSIRDLFADML-FGKSLSVNDSDSFIKINGSFTLITKNDNLNLLPNYHSLITKNVGYQI 757
Qy 245 LOSNLIAIVRCFWFTVESGLIENHEGRKAYGAVLISSPQELG----- 286
Db 758 VVNVFHIDARLLTAELQNTVFSNPK-----PVIKSPVELSKSLFEVWKTIFENSVNQI 810
Qy 287 ----HAFIDNVRVLPLEDOIIRLPNTSTPQETLFSIRHPD 324
Db 811 LKSYTFDNLKFFPKADGSSRLSDFLSDQDQVIFPAFVD 852

```

## RESULT 31

```

AAE11093
ID AAE11093 standard; peptide; 56 AA.
AC AAE11093;
XX
DT 18-DEC-2001 (first entry)
DE phhAB fusion peptide #4.
KW Phenylalanine hydroxylase; PAH; phhA; 4a-carbinolamine dehydratase; phhB;
KW aromatic aminotransferase; phhC; milk protein; animal protein; casein;
KW proteinaceous food product; globulin; whey protein; phenylketonuria;
KW PKU; inherited metabolic disorder; impaired brain function; neurotic;
KW cell therapy; phhAB fusion peptide.
XX Unidentified.
XX WO200168822-A2.
XX 20-SEP-2001.
XX 14-MAR-2001; 2001WO-DK00172.
XX 14-MAR-2000; 2000US-0525116.
XX (NITLA-) NILAB APS.
XX Johnsen M, Ravn P, Madsen SM, Vrang A, Israelsen H, Bredmose L;
XX Arnau J, Jensen SH, Gjetting T, Nielsen E;
XX WPI; 2001-590055/66.
XX Novel recombinant cells comprising a nucleic acid encoding a gene
XX product having phenylalanine hydroxylase activity, that is derived from
XX a prokaryotic organism, is useful for treating phenylketonuria in
XX mammals -
XX

```

PS Disclosure, Page 19; 91pp; English.

XX The patent discloses novel cells comprising a nucleic acid encoding  
 CC a gene product having phenylalanine hydroxylase (PAH) activity such  
 CC as phenylalanine hydroxylase (paha), 4a-carbinolamine dehydratase  
 CC (pnhb) and aromatic aminotransferase (pnhc), which are derived  
 CC from a prokaryotic organism. The patent also relates to fusion  
 CC proteins comprising a protein enhancing and/or stabilizing the  
 CC PAH activity in addition to PAH activity. The cells are useful  
 CC for preparing a proteinaceous food product (animal protein such  
 CC as a milk protein derived from casein, globulin or a whey protein)  
 CC having reduced content of phenylalanine. The method involves contacting  
 CC the food product starting material with the cells or fusion proteins  
 CC such that at least part of the phenylalanine content of the starting  
 CC material is converted into compounds that do not cause phenylketonuria  
 CC (PKU) by the enzymatically active product. PAH enzyme is useful for  
 CC manufacturing a medication for treating PKU, which is an inherited  
 CC metabolic disorder resulting in an accumulation in the body of  
 CC L-phenylalanine and metabolites that can cause impaired brain  
 CC function. The present sequence is a peptide sequence of pnhab  
 CC fusion protein.

XX Sequence 56 AA;

Query Match 5.2%; Score 98; DB 22; Length 56;  
 Best Local Similarity 35.7%; Pred. No. 0.024; 14; Indels 0; Gaps 0;  
 Matches 15; Conservative 13; Mismatches 14; Indels 0; Gaps 0;

QY 250 IAIYRCFWFTVESGLIENHGRKAYGAVLISPOELGHAFFID 291  
 DB 1 VFLARLYWMTIEFGLVETDQKRIYGGGILSSPKETVYSLSD 42

RESULT 32  
 ABB97206  
 ID ABB97206 standard; Protein; 689 AA.  
 XX ABB97206;  
 AC ABB97206;  
 XX 28-JUN-2002 (first entry)  
 DT 28-JUN-2002 (first entry)  
 DE Novel human protein SEQ ID NO: 474.  
 XX Human; anti-nausea; vulnerary; anti-inflammatory; immunomodulator;  
 KW anti-infectivity; cerebroprotective; cytoprotective; rheumatic; gene therapy;  
 KW neuroprotective; antiparkinsonian; protein therapy; EST;  
 KW expressed sequence tag.  
 XX Homo sapiens.  
 OS WO200222660-A2.  
 XX 21-MAR-2002.  
 PD 10-SEP-2001; 2001WO-US26015.  
 PF 11-SEP-2000; 2000US-0659671.  
 PR (HYSE-) HYSEQ INC.  
 PA Tang YT, Liu C, Zhou P, Aouni V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
 XX WPI: 2002-292408/33.  
 DR N-PSDB; ABR32392.  
 XX An isolated polynucleotide for treating diseases associated with its  
 PT encoded polypeptide such as cancer and multiple sclerosis -  
 PS Example 2; SEQ ID NO 474; 509pp; English.  
 CC The present invention provides the protein and coding sequences of 444

CC novel human proteins. These were isolated from expressed sequences tags  
 CC (ESTs). They can be used to stimulate cell growth, to regulate  
 CC hematopoiesis e.g. to treat aplastic anemia, to help tissue regrowth  
 CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
 CC multiple sclerosis, to regulate activity or inhibit e.g. to treat  
 CC infertility, to regulate haemostasis or thrombolysis e.g. to treat  
 CC stroke and cancer, to screen for drugs, to treat inflammatory conditions  
 CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.  
 CC Parkinson's disease. The present sequence is a protein of the invention.

XX Sequence 689 AA;

Query Match 5.2%; Score 98; DB 23; Length 689;  
 Best Local Similarity 22.2%; Pred. No. 0.9; 107; Indels 114; Gaps 17;  
 Matches 74; Conservative 38; Mismatches 107; Indels 114; Gaps 17;

QY 19 KLNQSLSLPFONS-----QSLQRAYSTPYSYRIILQENKQALARRHKCISILEFPKN 73  
 DB 188 KAAGSLLEFPDHLFTMLQELDKTPGESIAGYRICIOAIIQDKPKIAT-----AN 237  
 QY 74 L-FVHLLSLSKNREGCSTDW-AVSTPFNNMLWRLISPSFLW-----KSY 121  
 DB 238 LKQFLBELRSHQGRPAKCLTIWMALGQGFAN-----LLEGKWLGLMPLVLSIKSL 290  
 QY 122 CRRFEDLYLEAFGLSDPDLHQAVIKFPELETHPSYVSGFVAPHOYLSDDRYFPFA 181  
 DB 291 SP-FAIYYDLRLIM-----HNLTKGF-----GMIGKDPFPLDYAYMP-- 330  
 QY 182 SWKRTLDKNPFLTPDLIHDLGHVPM-----LHPSFSFPFIMGRLETKYI 229  
 DB 331 -----NNSLTPSLQQLQLYRKLKVLAFKAKDPSTLHTYPSFL----- 370  
 QY 230 EKVQALPS-----KKORIOTLQSLNIA--IVRCFTV-----ESGLNHEGRKAYGA 276  
 DB 371 --SRATPSCPPEKKELLSLRECLTVDPLASVWRQLYPRHLSQSLLEH----- 420  
 QY 277 VLISL-----POELGHAFFIDNVRVLPLELDQIIR 305  
 DB 421 LLSWEOIPIPKVQKSLQETIOSLKLITNOELIR 452

RESULT 33  
 AAY35368  
 ID AAY35368 standard; Protein; 439 AA.  
 XX AAY35368;  
 AC AAY35368;  
 XX 13-SEP-1999 (first entry)  
 DT 13-SEP-1999 (first entry)  
 DE C. pneumoniae protein involved in metabolism of nucleic acids.  
 XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope.  
 XX Chlamydia pneumoniae.  
 OS WO9927105-A2.  
 XX 03-JUN-1999.  
 PD 20-NOV-1998; 98WO-IB01890.  
 PF 04-NOV-1998; 98US-0107078.  
 PR 21-NOV-1997; 97FR-0014673.  
 XX (GEST ) GENSET.  
 PA Griffais R;  
 PI WPI: 1999-357842/30.  
 XX Genome sequence of Chlamydia pneumoniae

[illegible]





[illegible]

```
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160988.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161407.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.0%; Score 95; DB 21; Length 944;
Best Local Similarity 21.7%; Pred. No. 2.8;
Matches 97; Conservative 48; Mismatches 128; Indels 174; Gaps 23;
```

```
QY 9 DPYIILKIALKROSILFQNSOSLQRA-----YSTPYSYRIILQENKKEQALAR 61
DB 395 NPAIVQKVVDELSPSNFRIFWMSQKFEQGTDKAPMYNTAVSLEKITSS----- 443
QY 62 HKCISILFEFFKNLIFVHL-----LSLSKQREGSCTDMVVSSTPFNNRLMYR 109
DB 444 ----TQEWVQSPVHLHLPAENVIFPTDSL-KDADDEKYTPVLLRKTFF--SRLMYK 496
QY 110 LLSRFSLMKSY-----CP-----RPFIDYLEAF----- 133
DB 497 -PDMFSKPKAYKMDPNCPLAVSSPDAAVLTDTFTRLMDYLANEYAYVAQVAGLYYGS 555
QY 134 -----GLISDF--LDHOAVIKFPELTFHESYYPVSGFVA 165
DB 556 LSNNGFELTLGYNHKLRIILFTVVGKIANEVXPDPAVAK--ETVTKQYKFKRQ 611
QY 166 P-HQ---YLSL-LQDRYFPIASVMTLTDKNFSILTPDLIHDLGH-----VFWLLH 211
```

```
DB 612 PYHQAMTCGLIIQDOTWP-----WTEEL--DVLSHLEAEVAKFVPMILS 655
QY 212 PSFSEFPF-----NMGRLEFTVIERV-----QALPSKKORIQTLSMLIAIVRC 255
DB 656 RTFIECYIAQNVNNEAESMVGHIEDVLFNDPKPICRPFPS-----QHLTRVVKL--- 707
QY 256 FWFVESGLIENHGRKAYGAVALISSPQELGHAFIDNVRVLPLELQIIRLP-FNTSTPQ 314
DB 708 -----GEGMKYFYHQDGSNPSDENSAHVHYIQVHRDPSMNKIKQLFLGLVAKQ 755
QY 315 ETLFSIRHDELVELTSKLEMMIDQGL 341
DB 756 ATFHQLRVQLGYITA-LACRNDSGI 781
```

## RESULT 36

AAG31677 standard; Protein; 970 AA.

AAG31677;

17-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 38082.

KW Protein identification; signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.

Arabidopsis thaliana.

EPI033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

```
XX 25-FEB-1999; 99US-01231825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0128845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 07-MAY-1999; 99US-0132486.
XX 11-MAY-1999; 99US-0132863.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134841.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 28-MAY-1999; 99US-0136392.
XX 01-JUN-1999; 99US-0137222.
```

PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137503.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139453.	PR	12-AUG-1999;	99US-0148545.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139454.	PR	13-AUG-1999;	99US-0149358.
PR	18-JUN-1999;	99US-0139455.	PR	16-AUG-1999;	99US-0149317.
PR	18-JUN-1999;	99US-0139456.	PR	17-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	18-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	20-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	23-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	25-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	26-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	27-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140354.	PR	31-AUG-1999;	99US-0151438.
PR	24-JUN-1999;	99US-0140695.	PR	01-SEP-1999;	99US-0151930.
PR	28-JUN-1999;	99US-0140623.	PR	07-SEP-1999;	99US-0152363.
PR	29-JUN-1999;	99US-0140991.	PR	10-SEP-1999;	99US-0153070.
PR	30-JUN-1999;	99US-0141287.	PR	13-SEP-1999;	99US-0153758.
PR	01-JUL-1999;	99US-0141842.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0142154.	PR	16-SEP-1999;	99US-0154039.
PR	02-JUL-1999;	99US-0142055.	PR	20-SEP-1999;	99US-0154779.
PR	06-JUL-1999;	99US-0142390.	PR	22-SEP-1999;	99US-0155139.
PR	08-JUL-1999;	99US-0142803.	PR	23-SEP-1999;	99US-0155486.
PR	09-JUL-1999;	99US-0142920.	PR	24-SEP-1999;	99US-0155659.
PR	12-JUL-1999;	99US-0142977.	PR	28-SEP-1999;	99US-0156458.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	07-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	08-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144684.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160981.
PR	26-JUL-1999;	99US-0145276.	PR	22-OCT-1999;	99US-0160989.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161406.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161359.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161320.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161392.
PR	04-AUG-1999;	99US-0147204.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147302.	PR	29-OCT-1999;	99US-0162142.
PR	05-AUG-1999;	99US-0147192.			

Query Match

5.0%; Score 95; DB 21; Length 970;

Best Local Similarity 21.7%; Pred. No. 3;  
Matches 97; Conservative 48; Mismatches 128; Indels 174; Gaps 23;

```

QY 9 DPKYIKALKIKRQSLFLFQNSQSLQRA-----YSTYSYTRIILOKENEKQALAR 61
DB 421 NPAIVOKVVDLSPNSFRIFMBSQFSDTKABPMYNIVASLEKITSS----- 469
QY 62 HKCISLIEFFKNLLFVHL-----LSLKNORBCSCSDMAVSTPPNNMYR 109
DB 470 ----TIOEWQASAPVHLHLPAVNFILPTDLSL-KDADDEKETVPVLRKTPF--SRLMYK 522
QY 110 LLSRPSLWKSYP-----CP-----RFLDYLEAF----- 133
DB 523 -PDTMFSKPKAVKMDPNCPLAVSSPDAAVLTIDTFLMDYNEVAYVAQVAGLYGVS 581
QY 134 -----GLISDF--LDHQAVIKFEFLETHFSYYPVSGFVA 165
DB 582 LSDNGPELTLGLVNHKRLILETVVGKIANFEVAPRFAVIK---ETVTKKQYKFRQ 637
QY 166 P-HQ---YLSL-LQDRYPPIASVMTDKNFSLTPDLIDLGH-----VPMILH 211
DB 638 PYHQAMYCSLILQDOTP-----WTEBL--DVLSHLEADVAKFPMILS 681
QY 212 PSFSEFFI-----NMGRIFTYVIEKV-----QALPSKKORIOTLOSNIATVRC 255
DB 682 RTIECTINGNVENNEAESVYKHIEDVLFNDPKPICRPLPS-----QHLTKRVKL--- 733
QY 256 FWFVESGLIENHGRKAYGAVLISSPOELGHAFTDNVRYLPLEDOITRLP-FNTSTPQ 314
DB 734 -----GEGMKKFYHODGSPSDENSALVHYIQVHRDPSMNTIKLOLFGVLAKQ 781
QY 315 ETLFSIRHFDLVELTSLKLEWMLDGL 341
DB 782 ATFHQRTVEQLGYITA-LAQRNDGI 807

RESULT 37
AAM55671
ID AAM55671 standard; Protein; 293 AA.
XX
XX AAM55671;
XX
DT 24-JUN-1998 (first entry)
XX
DE H. pylori ORF hp5p15641_5211687_c2_29 cellular protein.
XX
XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
XX identification; binding compound; bactericidal; life cycle; activator;
XX inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
XX
XX Helicobacter pylori.
XX
XX WO9737044-A1.
XX
XX 09-OCT-1997.
XX
XX 27-MAR-1997; 97WO-US05223.
XX
XX 06-DEC-1996; 96US-0761318.
XX 29-MAR-1996; 96US-0625811.
XX 02-APR-1996; 96US-0758731.
XX 25-OCT-1996; 96US-0736905.
XX 28-OCT-1996; 96US-0738859.
XX
XX (ASTR ) ASTRA AB.
XX
XX Alm RA, Smith D;
XX
XX WPI, 1997-503122/46.
XX
XX N-PSDB; AAV25080.
XX
XX Helicobacter pylori nucleic acid sequences and encoded
XX polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT

```

PT infection and for diagnosis of H. pylori infection

PS Claims 14,93; Pages 878-879; 1145pp; English.

XX This is the sequence of a Helicobacter pylori cellular protein.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds. The  
 CC use as potential H. pylori life cycle activators or inhibitors. The  
 CC DNA and probes derived from it may be used for the identification of  
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic  
 CC acid sequences complementary to the DNA act as antisense sequences and  
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies  
 CC against the protein can be used in immunoassays to evaluate the abundance  
 CC and distribution of H. pylori-specific antigens. The genomic sequence of  
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated  
 CC by mechanically shearing the bacterial DNA. The sequences were analysed  
 CC for ORF of at least 180 nucleotides, and the predicted coding regions  
 CC defined by computer evaluation. To identify likely H. pylori antigens for  
 CC vaccine development, the amino acid sequences predicted from various ORF  
 CC were analysed for significant homology to other known or exported  
 CC membrane proteins. Having identified and determined the sequences of  
 CC interest, particular regions can be isolated from H. pylori by PCR  
 CC amplification for recombinant polypeptide production, e.g. in E. coli  
 CC hosts.

Sequence 293 AA;

Query Match 5.0%; Score 94; DB 18; Length 293;

Best Local Similarity 21.0%; Pred. No. 0.67; 67; Indels 38; Gaps 6;

Matches 38; Conservative 38; Mismatches 38; Indels 38; Gaps 6;

```

QY 179 PIASVMTLKDKNFSLTPDLIDLGHVP-----W-----LIH 211
DB 93 PIAMTAQK-BEDVLKLSVLELIDLTGKAPGYVAPWMEFSNITNELLKRGFYDHSLMH 151
QY 212 PSFSEFFINMGRLFTYVIEKVQALPSKKORIOTLOSNIATVRCFWFVESGLIENHGR 271
DB 152 NDFTPYFVRVQDSMKIDSLKAKMKMKPLIRGYETNIVELPAMWYLDLPPMKFKSP 211
QY 272 KAYGAVLISSPOELGHAFTDNVRYLPLEDOITRLP---PNTSTPOETLFS---IRHPD 324
DB 212 NSFQV---SPRDIQGMWIDQFDWYVREMDYAVFSMTIHDPVARSAPVLLMHKEIIEHIN 268
QY 325 E 325
DB 269 Q 269

RESULT 38
AAG82418
ID AAG82418 standard; Protein; 528 AA.
XX
XX AAG82418;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO.1930.
XX
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX vaccination; endocarditis.
XX
XX Staphylococcus epidermidis.
XX
XX WO200134809-A2.
XX
XX 17-MAY-2001.
XX
XX 09-NOV-2000; 2000MO-US30782.
XX
XX 09-NOV-1999; 99US-0164258.
XX
XX (GLAX ) GLAXO GROUP LTD.
XX

```



QY 209 LHP--FSEFFINMGRLEFKVIEKQALPSKQRIQTQSNL--IAIVRCFWFTVESGL 264  
 Db 297 SIHPSRQFDHYESN-----QVAKLLLL-----RINSTDDEHFEQIDAIMSLVQ-L 344  
 QY 265 IENHEGRK-----AYGAV-----LISPP-----OELGAFID--- 291  
 Db 345 AKVFSACRQKRVSVMMPSLYRLNLITLITGIIKVPYIRKLISGGLLYLDSDNDKVFVQL 404  
 QY 292 --NVRVLPLELDQIIRLEFPNTSPQETLFSIRHFDDELVELTSKL----- 333  
 Db 405 LINDKISPLMKSQYNNVLRNMEYDKVELEFNFQDQVLEITBQIKRILISNDITNLQLSK 464  
 QY 334 -----EWMLD-----QGLIESPLYNQEKYISGFEVLC 361  
 Db 465 TPLSIKIMVAMWYLSHLCSGLISV--NRIVLKLKIRKIFC 502

RESULT 40  
 ABB55057  
 ID ABB55057 standard; Protein; 307 AA.

XX AC ABB55057;  
 XX DT 16-MAY-2002 (first entry)  
 XX DE Lactococcus lactis protein prnd.  
 XX KM Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.  
 XX OS Lactococcus lactis IL1403.  
 XX PN FR2807446-A1.  
 XX PD 12-OCT-2001.  
 XX PF 11-APR-2000; 2000FR-0004630.  
 XX PR 11-APR-2000; 2000FR-0004630.  
 XX PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.  
 XX PI Bolocine A, Sorokine A, Renault P, Ehrlich SD,  
 XX DR WPI; 2002-043418/06.  
 XX PT New nucleotide sequence useful in the identification or Lactococcus  
 XX PS lactis and related species -  
 XX Claim 6; SEQ ID No 1759; 2504bp; French.  
 XX The present invention is related to a Lactococcus lactis nucleotide  
 CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The  
 CC nucleic acid sequence is useful in the detection and/or amplification of  
 CC nucleic acid sequence, particularly to identify Lactococcus lactis or  
 CC related species. The proteins of the invention are useful for the  
 CC biosynthesis or biodegradation of a composition of interest. The  
 CC invention helps research in lactic bacteria, particularly useful in the  
 CC production of yogurt and cheese.  
 CC Note: The sequence data for this patent is based on equivalent patent  
 CC WO200177334 (published 18-OCT-2001) which is available in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 307 AA;

Query Match 4.9%; Score 93; DB 23; Length 307;  
 Best Local Similarity 20.9%; Pred. No. 0.9; 93; Indels 122; Gaps 16;  
 Matches 67; Conservative 38; Mismatches 93;

QY 15 KIAL--KLRQST---SLPFQNSQISQRAYSTPYRYRIILQKE-----NKEQALARHKC 64  
 Db 5 KVTLDKKIRSYVMRSMFLQSGMYERNQNGMAYSLIPALKKLYPSGEBAKALKRH-- 62

QY 65 ISLEPFKNLLFYH-----LISLKNQREGCSTD-----NAVSTPEFNR 104  
 Db 63 ---LEFNTHPYVAADIIQVTLALBEBRANGADIDDAIQGVKVMGPIAGIGDVF-- 117  
 QY 105 NLWY-----RLSSRFSIMKSCYCRF-----FLDYLEAF----- 133  
 Db 118 --MFTVRPIVGAIALAATGAGSIIAPLFFPIVMNAIRIAFIWYQEFYKSGSAITKDG 175  
 QY 134 -GLISDFLDQAVIKFEELETHPSYYPVSGFVAPHOYS---LTDREY--FPJASYMRTL 187  
 Db 176 GGLLQGTATKASITLGMFVGLIQHWITNFNGPAAVSKPLQKGAYLEFPFGSVSGT- 234  
 QY 168 DKDNPSLTPDLIDILGHVPMWLMHSFSSFFINMGRLEFKVIEKQALPSKQRIQTQSNL 247  
 Db 235 -----QHDILQVQ-----NKLSDPTK--VYVLD 259  
 QY 248 NLIAIV-----RCFW 257  
 Db 260 NLNDQILPGLAGLLITPLCWM 279

RESULT 41  
 ABG15643  
 ID ABG15643 standard; Protein; 744 AA.

XX AC ABG15643;  
 XX DT 18-FEB-2002 (first entry)  
 XX DE Novel human diagnostic protein.#15634.  
 XX KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX OS Homo sapiens.  
 XX PN WO200175067-A2.  
 XX PD 11-OCT-2001.  
 XX PF 30-MAR-2001; 2001MO-US08631.  
 XX PR 31-MAR-2000; 2000US-0540217.  
 XX PR 23-AUG-2000; 2000US-0649167.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Drmanac RT, Liu C, Tang YT;  
 XX DR N-PSDB; AAS79830.  
 XX WPI; 2001-639362/73.  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue; as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations in

Claim 20; SEQ ID No 46002; 103bp; English.

CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG030177 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 744 AA;

Query Match 4.9%; Score 93; DB 22; Length 744;  
 Best Local Similarity 19.2%; Pred. No. 3.2; Indels 152; Gaps 22;  
 Matches 87; Conservative 73; Mismatches 152; Indels 152; Gaps 22;  
 QY 3 YCERLDPKYLKALKLQSLSLFQNSQSLORAYSTPSYY-RIILOKENEKQAL-- 59  
 DB 335 YAPNTGAPRFKQVLSLQDL----DSHTLIMGDFNPLSLDRSMRQVKNKDTQELNS 390  
 QY 60 ARKICISILEFFKVL-----LFVHLLS-----LSKNQR-----EGCSTD 93  
 DB 391 ALHOA-PLIDIYRTLHPKSTAYTFFSAPHHTYSKIDHILGSKALLSKCRTEITITNCLSD 449  
 QY 94 MAVVSTPFPNRLMYRLLSRFSLWKSYPREFLDYLEAFGLLSDFLDH---QAVIK-FF 149  
 DB 450 HSAIKLELRKKN---LTQNRSTTKW-----LNNL-----LNDYVWVHNMKAEIKMFF 494  
 QY 150 ELE-----THFSYYPVSGFVAPHQYLSL-----LDQRYPIASVMRTLDKDNFSLTPD 197  
 DB 495 ETNENKDTITQNLWDAPKACVCRKPTALNAHKRQKERSKIDTLTSQLEKEQKIN--- 550  
 QY 198 LIHDLGHVPLWLPSPSEFFINMGLFTKVEK-----VOALPSKKQRIQTLSQNLIAI 252  
 DB 551 -----KIDPLARLIKKEKNQIDAINKDKGIDITDPTIKTT 588  
 QY 253 VRCFWFTVESGLIENH-----GRKAYGAVLS-----S 281  
 DB 589 IREYKGLHYANKLENLEMDKFDKYLPRLNQBEVESLNRPITGAEIVAINSLPTKKS 648  
 QY 282 P-----OELGHAFIDNVRVLPLELDQIIRLPNTSTPQETL-----FSIRHFDVLVELTS 331  
 DB 649 PGPDGFTAEFYQYKEALLVCPFTLFE--KHQVNYDFPQETLHFSPLSRSSBELNSTY 706  
 QY 332 KL-----EMWLDQGLLESIPLYNQ 350  
 DB 707 RLYGWSEBQSLWIRTWARVQKPRAAVPLVNE 738

RESULT 42

AAU96092

ID AAU96092 standard; Protein; 333 AA.

XX AC AAU96092;

XX AC AAU96092;

XX 15-JUL-2002 (first entry)

XX RhCMV unique short region US28.4 protein.

XX CMV; US28; vaccine; virucide; neuroprotective;

XX antiinflammatory; ophthalmological; retinitis; encephalitis;

XX pneumocystis.

XX Rhesus cytomegalovirus.

XX WO200218954-A2.

XX 07-MAR-2002.

XX 30-AUG-2001; 2001WO-US27392.

XX 30-AUG-2000; 2000US-229365P.

XX (CHEM-) CHEMOCENTRYX INC.

PI Schall TJ, Penfold M;

XX WPI; 2002-351718/38.

DR N-PSDB; ABK68823.

XX Isolated or recombinant homologues of US28 proteins and nucleic acids

PT encoding the proteins, for use in vaccine compositions for treating an

PT animal infected with, or at risk of infection by, cytomegalovirus -

XX Claim 69; Page 67; 95pp; English.

XX The invention relates to an isolated or recombinant homologue of US28  
 CC protein which binds a chemokine (encoded by an open reading frame in the  
 CC unique short (US) region in human cytomegalovirus (CMV) genome, having  
 CC at least 75% identity to a rhesus monkey RHUS28.1-RHUS28.5; RH78 and  
 CC RHUS33 (encoded by an open reading frame 78 in the unique long (UL) and  
 CC ORF 33 is the US region of CMV genome) and their encoding nucleic acids.  
 CC Also included are a vector comprising the nucleic acids, a cell  
 CC comprising the nucleic acids, a vaccine comprising an immunogenic CMV  
 CC polypeptide encoded by at least a region of CMV genome in which the  
 CC polynucleotide sequence encoding US28 or its homologue has been  
 CC inactivated, identifying an agent that reduces CMV dissemination in an  
 CC animal, by determining whether the agent inhibits the expression or  
 CC activity of US28 or US28 homologue, or fragment or variant of US28 or  
 CC US28 homologue and treating an animal infected with CMV or at risk of  
 CC infection by CMV, by administering the agent to the animal. The vaccine  
 CC is useful for treating an animal infected with cytomegalovirus or at a  
 CC risk of infection by CMV. The vaccine is also useful for inducing a  
 CC therapeutic or protective immune response in a patient and in methods for  
 CC treating diseases including retinitis, encephalitis and pneumocystis  
 CC caused by CMV infection. The present sequence represents a human CMV or  
 CC monkey CMV US or UL protein used to construct the vaccine of the  
 CC invention.

XX Sequence 333 AA;

Query Match 4.9%; Score 92; DB 23; Length 333;

Best Local Similarity 21.7%; Pred. No. 1.3;

Matches 60; Conservative 44; Mismatches 101; Indels 72; Gaps 12;

QY 61 RHKICISILEFPKLLFVHLLSLSKNQRQECSTDMAVVSTPFPNRLMYRLLSRFSLWKS 120

DB 50 RHRTNS---FSDVLFPHLM-----ITBEVFTLTP-----VWAYHLTTHGNLPGS 91

QY 121 YCPFRF-FUDYLEAFGLLSDFLDHQAIVKPFLETHFSYYPVSGFVAPHQYLSLQDRYFP 179

DB 92 WCRSLTFVFLTVFAR-----AFF-----YLLLIWDYSV 121

QY 180 IASVMRTLDKDNFSLTPDLIHDILGHVPMWLHP-SFSEFFINMGLRFTKVIKQVQALPSK 238

DB 122 IICRHPLPVNLNYS-----QVIGLSVWLVAVLSASPFSIFNGSV-KQCLGNMGSIPE 173

QY 239 KQRIOTLOSNIATVRCFWFTVESGLIENHGRKAYGAVLSQPELGHAFIDNVRVLP 298

DB 174 SSAVLNLEVLCS---FWLPLIM-----SANCYQAKRRASPDQLHLYRCSLITIT 223

QY 299 ELDQIIRLPNTSTPQETLFSIRHFDVLVELTSKLEW 335

DB 224 TTYAIVMFPFPHALLIDALISISH-----VEPSSALHW 256

RESULT 43

AAU33268

ID AAU33268 standard; Protein; 919 AA.

XX AC AAU33268;

XX AC AAU33268;

XX 18-DEC-2001 (first entry)

XX Novel human secreted protein #3759.

XX Human; vaccination; gene therapy; nutritional supplement;

XX stem cell proliferation; haematopoiesis; nerve tissue regeneration;

KM immune suppression; immune stimulation; anti-inflammatory; leukaemia.  
 XX Homo sapiens.  
 OS  
 XX  
 PN WO200179449-A2.  
 XX  
 XX  
 PD 25-OCT-2001.  
 XX  
 XX 16-APR-2001; 2001WO-US08656.  
 XX  
 XX 18-APR-2000; 2000US-0552928.  
 PR 26-JAN-2001; 2001US-0770160.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 PA  
 PI Tang YT, Liu C, Drmanac RT;  
 PI  
 DR WPI; 2001-611725/70.  
 XX  
 XX  
 PT Nucleic acids encoding a range of human polypeptides, useful in genetic  
 PT vaccination, testing and therapy -  
 PS  
 PS Claim 20; Page 750; 765pp; English.  
 XX  
 CC The invention relates to novel human secreted polypeptides. The  
 CC polypeptides and antibodies to the polypeptides are useful for  
 CC determining the presence of or predisposition to a disease associated  
 CC with altered levels of polypeptide. The polypeptides are also useful for  
 CC identifying agents (agonists and antagonists) that bind to them. Cells  
 CC expressing the proteins are useful for identifying a therapeutic agent  
 CC for use in treatment of a pathology related to aberrant expression or  
 CC physiological interactions of the polypeptide. Vectors comprising  
 CC the nucleic acids encoding the polypeptides and cells genetically  
 CC engineered to express them are also useful for producing the proteins.  
 CC The proteins are useful in genetic vaccination, testing and  
 CC therapy, and can be used as nutritional supplements. They may be used to  
 CC increase stem cell proliferation; to regulate hematopoiesis; and in  
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;  
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and  
 CC in treatment of leukaemia. AAU29510-AAU33304 represent the amino acid  
 CC sequences of novel human secreted proteins of the invention.  
 CC  
 XX  
 XX  
 SO Sequence 919 AA;  
 Query Match 4.9%; Score 92; DB 22; Length 919;  
 Best Local Similarity 20.7%; Pred. No. 5.5; Indels 68; Gaps 12;  
 Matches 61; Conservative 48; Mismatches 118;  
 QY 3 YCERLDPKYLKIALKRLSLSLFPONSOLQRAYSTPYSYR---IIQKENEKQAL 59  
 DB 567 YTRBELCTMFIREVGTLSQVSKVNGSEIL-----FSYQDLVITLPFE-----L 613  
 QY 60 ARHKCISILEFFKRLLFVHLISLKNQREGCSTMAVVSFPFRNLW-----YRLISS 113  
 DB 614 RKRLIDVISMYREL-----LKOLSKAQVFPKAIQSLKTEVL-RNLDLLQPIQLIED 668  
 QY 114 RFLMKSCYCRFFLDYL--EAFGLSDFLDQAVIKFELETHSYYPVSGFAP----- 166  
 DB 669 NIKOLKEMKFTYLIINVIDEINTITFNDYIPY--VFKLKLNCLINLHKNEFIQNELQEA 726  
 QY 167 -----HOYLSLQDRYFPFASVMT-----LDKDNFSLTPDLI----- 199  
 DB 727 SQELQIHQYIMALREEVDPDSIVGWTKYVELEEKIVSLIKNLVALADPSEYIVSAS 786  
 QY 200 ---HDLGHVPMILHPSSEFPINNGRLFTVIEKVOALSKKQRIQTQSNLIA 251  
 DB 787 NFTSQLSQVBOGFLHNTQEVISLITDPDGKXKIKMLSATQOEI--IKSQALIA 839  
 RESULT 44  
 AAU31237  
 ID AAU31237 standard; Protein; 3923 AA.

AC AAU31237;  
 XX  
 XX 08-NOV-1999 (first entry)  
 DT  
 XX  
 XX Human Apo B protein fragment.  
 DE  
 XX  
 XX Apo B; hyperlipidemia; human; treatment; hepatocyte; apoprotein;  
 KM Apo A1; low density lipoprotein; LDL; blood; therapy; atherosclerosis;  
 KM high density lipoprotein; HDL; cholesterol; coronary heart disease;  
 KM Alzheimer's disease; hypobetalipoproteinemia; dybetalipoproteinemia.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT 1..3923  
 FT Protein /label= Apo B  
 FT /note= "partial sequence, no start codon given"  
 XX  
 XX  
 PN WO9940789-A1.  
 XX  
 XX 19-AUG-1999.  
 PD  
 XX  
 XX 28-AUG-1998; 98WO-US17908.  
 PF  
 XX 30-JUN-1998; 98US-0108006.  
 PR 12-FEB-1998; 98US-0074497.  
 XX  
 XX (MINU ) UNIV MINNESOTA.  
 PA (YESH ) UNIV YESHIVA EINSTEIN COLLEGE.  
 PA  
 PI Bandopadhyay PT, Kren BT, Roy-Chowdhury J, Steer CJ;  
 PI  
 DR WPI; 1999-527333/44.  
 DR N-PSDB; AAZ09525.  
 XX  
 XX Mutating apolipoprotein genes in hepatocytes to control cholesterol  
 PT levels, e.g. for treating or preventing hyperlipidemia, particularly  
 PT atherosclerosis  
 PT  
 PS Claim 4; Page 75-83; 106pp; English.  
 XX  
 XX This invention describes a novel method for the genetic treatment of  
 CC hyperlipidemia by altering genes, in hepatocytes, for apoprotein (apo) B,  
 CC E or A1. Low density lipoprotein (LDL) levels in the blood are reduced by  
 CC altering an apo B gene (1) in a hepatocyte. The invention describes a  
 CC method for the therapeutic and/or prophylactic method involving altering  
 CC an apo B gene in hepatocytes by introducing the mutations Arg120Ys,  
 CC Arg156Gys or Cys158Arg and a method for ameliorating atherosclerosis by  
 CC altering the apo A1 gene in a hepatocyte so that the altered protein can  
 CC dimerize. Altering expression of apo genes regulates levels of high and  
 CC low density lipoprotein cholesterol. Altering expression of apo B, E and  
 CC A1 genes is used to treat or prevent atherosclerosis, coronary heart  
 CC disease, Alzheimer's disease, hypobetalipoproteinemia, and  
 CC dybetalipoproteinemia. This sequence represents a fragment of the human  
 CC Apo B protein described in the method of the invention.  
 CC  
 XX  
 XX  
 SO Sequence 3923 AA;  
 Query Match 4.9%; Score 92; DB 20; Length 3923;  
 Best Local Similarity 20.7%; Pred. No. 44;  
 Matches 61; Conservative 48; Mismatches 118; Indels 68; Gaps 12;  
 QY 3 YCERLDPKYLKIALKRLSLSLFPONSOLQRAYSTPYSYR---IIQKENEKQAL 59  
 DB 3571 YTRBELCTMFIREVGTLSQVSKVNGSEIL-----FSYQDLVITLPFE-----L 3617  
 QY 60 ARHKCISILEFFKRLLFVHLISLKNQREGCSTMAVVSFPFRNLW-----YRLISS 113  
 DB 3618 RKRLIDVISMYREL-----LKOLSKAQVFPKAIQSLKTEVL-RNLDLLQPIQLIED 3672  
 QY 114 RFLMKSCYCRFFLDYL--EAFGLSDFLDQAVIKFELETHSYYPVSGFAP----- 166  
 DB 3673 NIKOLKEMKFTYLIINVIDEINTITFNDYIPY--VFKLKLNCLINLHKNEFIQNELQEA 3730



QY 167 -----HOYLSLQDRYFPFIASVMRT-----LDKNFSLTPDLI----- 199  
 DB 3731 SOELQOIHOYIMALREYFDPISVGTWTVKYVELEEKIVSLIKNLLVALKDFHSEYIVSAS 3790  
 QY 200 ---HDLGHVPLWLPSPFSFFINMGRLEFTKVIKQVQALPSKKQRIOTLQSNLIA 251  
 DB 3791 NFTSOLSSQVEQFLHRNIQIYLSILTDPDGKGEKIEIATSATQEI--IKSQAIA 3843

RESULT 45  
 AAW41262  
 ID AAW41262 standard; peptide; 4536 AA.  
 AC AAW41262;  
 DT 19-MAY-1998 (first entry)  
 DE Apolipoprotein B-100.  
 KW Anti-coagulant; apolipoprotein B-100; apoB-100; metastatic spread;  
 KW thromboplastin-mediated process; cancer; inhibitor; blood coagulation;  
 KW angiogenesis; cellular differentiation; apoptosis; KRAB-14;  
 KW prothrombinase complex.  
 OS Homo sapiens.  
 PN WO9743311-A1.  
 PD 20-NOV-1997.  
 PF 09-MAY-1997; 97WO-GB01255.  
 PR 09-MAY-1996; 96GB-0009702.  
 XX (UNLO ) ROYAL FREE HOSPITAL SCHOOL MED.  
 PA Bruckdorfer KR, Bettelaie C;  
 PI WPI, 1998-008798/01.  
 DR  
 XX  
 PT Peptide fragments of apolipo:protein B-100 with anticoagulant  
 PT activity - used for treating or preventing coagulation, inhibiting  
 PT angiogenesis, cell differentiation and apoptosis  
 XX  
 PS Disclosure; Page 42-47; 60pp; English.  
 XX  
 CC This sequence represents the human apolipoprotein B-100 (apoB-100).  
 CC Fragments of this sequence can be used in the peptide of the invention,  
 CC which has the formula (I), or their variants with one or more internal  
 CC deletions, insertions or substitutions, while retaining anti-coagulant  
 CC properties of apolipoprotein B-100 (apoB-100).  
 CC Z1-KAQ-X1-KKNKRRHS-X2-T-22 (I); X1 = S or Y; X2 = T or I;  
 CC Z1 = the N terminus of the peptide, or 1-47 amino acids (aa);  
 CC Z2 = the C terminus of the peptide, a terminal amide group or 1-77 aa.  
 CC Compositions containing the peptide are used for simultaneous, separate  
 CC or sequential treatment of cancer, particularly to prevent metastatic  
 CC spread. They are also used to inhibit thromboplastin-mediated processes,  
 CC specifically to prevent or reduce blood coagulation (e.g. during or after  
 CC surgery or in cases of heart attack, stroke etc.) and to inhibit  
 CC angiogenesis, cellular differentiation or apoptosis. KRAB-14, which is  
 CC active as such or as part of a 98-aa peptide, inhibits activation of the  
 CC prothrombinase complex; and prevents activation of factor VII on the  
 CC surface of thromboplastin and of platelets by thrombin. It binds to the  
 CC residues 58-66 of thromboplastin. Since (I) are much smaller than  
 CC apoB-100, they act more quickly.  
 XX  
 SQ Sequence 4536 AA;  
 Query Match 4.9%; Score 92; DB 19; Length 4536;  
 Best Local Similarity 20.7%; Pred. NO. 55;  
 Matches 61; Conservative 48; Mismatches 118; Indels 68; Gaps 12;

QY 3 YCERLDPKYLKIALKRLQSLFFQNSQSLQRAYSTPYSYR---IILOKENKEKQAL 59  
 DB 4184 YTBRELCTMTFREVTGTVLSQVSKVHNGSEIL-----FSYFQDLVITLPFE-----L 4230  
 QY 60 ARHKCISILEFFKNLLFVHLLSLSKNQREGCTDMVVSTPFFNRNLW-----YRLSS 113  
 DB 4231 RKHLIDVISMYREL-----LKDLSEKAEQSVFKAIQSLKTTTEVL--RNLQDLLQIFQIED 4285  
 QY 114 RPSLWKSYPFRFLDYL--EAFGLLSDFLDHQAVIKFFFELETHFSYYPVSGFVAP----- 166  
 DB 4286 NIKQLKEMKFTYLINYIQDEINTIFNDYIPY--VFKLLKENLCLNLHKNEFIQNELQEA 4343  
 QY 167 -----HOYLSLQDRYFPFIASVMRT-----LDKNFSLTPDLI----- 199  
 DB 4344 SOELQOIHOYIMALREYFDPISVGTWTVKYVELEEKIVSLIKNLLVALKDFHSEYIVSAS 4403  
 QY 200 ---HDLGHVPLWLPSPFSFFINMGRLEFTKVIKQVQALPSKKQRIOTLQSNLIA 251  
 DB 4404 NFTSOLSSQVEQFLHRNIQIYLSILTDPDGKGEKIEIATSATQEI--IKSQAIA 4456

RESULT 46  
 AAW96826  
 ID AAW96826 standard; protein; 4536 AA.  
 AC AAW96826;  
 DT 22-APR-1999 (first entry)  
 DE Amino acid sequence of human apolipoprotein B-100 (apoB-100).  
 KW Human apolipoprotein B-100; apoB-100; very-low density lipoprotein; VLDL;  
 KW apolipoprotein; binding; in vivo transport; nucleic acid; binding domain;  
 KW nuclear localisation sequence; gene therapy; cancer; cystic fibrosis;  
 KW non-small cell lung carcinoma; diabetes; arteriosclerosis.  
 OS Homo sapiens.  
 PN WO9856938-A1.  
 PD 17-DEC-1998.  
 PF 10-JUN-1998; 98WO-US11927.  
 PR 14-MAY-1998; 98US-0079030.  
 PR 13-JUN-1997; 97US-0874807.  
 XX (BAYU ) BAYLOR COLLEGE MEDICINE.  
 PA Guevara JG, Hoogvee RC, Moore JP;  
 PI WPI, 1999-070331/06.  
 DR  
 XX  
 PT Composition comprising nucleic acid bound to LDL or VLDL lipoprotein  
 PT - used for delivering nucleic acid to cells for gene therapy and  
 PT antisense treatment  
 XX  
 PS Disclosure; Fig 1A-C; 293pp; English.  
 CC  
 CC The present sequence represents human apolipoprotein B-100 (apoB-100).  
 CC ApoB-100 is a major apoprotein component of very-low density  
 CC lipoproteins (VLDL), intermediate density lipoprotein (IDL), low  
 CC density lipoproteins (LDL) and lipoprotein a. The specification  
 CC describes a composition that comprises LDL and apolipoproteins for the  
 CC binding and in vivo transport of nucleic acids. Binding domains  
 CC (see AAW96827-77) and nuclear localisation sequences (see AAW96878-97)  
 CC for use in the composition can be derived from the present sequence. The  
 CC composition is used to deliver nucleic acids to eukaryotic cells, in  
 CC vivo or in vitro, for expressing a therapeutic polypeptide or antisense  
 CC molecule (or ribozyme). Specifically they are used for gene therapy of  
 CC cancers (particularly non-small cell lung carcinoma), diabetes, cystic  
 CC fibrosis and arteriosclerosis.  
 XX



XX WPI; 1999-562049/47.  
DR N-PSDB; AAZ19859.  
XX  
PT New DNA polymerase from Thermoanaerobacter thermohydrosulfuricus, used  
PT for DNA amplification, sequencing, polymerase chain reaction and  
PT reverse transcription  
XX

XX Example 6; Fig 2; 71pp; English.

XX This sequence represents the DNA polymerase of Thermoanaerobacter  
CC thermohydrosulfuricus (Tt), deduced from PCR amplified genomic  
CC DNA (see AAZ19859). Claimed DNA polymerases (see AAY31813, AAY31815 and  
CC AAY31816) have the exonuclease activity of this enzyme removed by an  
CC having 540-582 amino acids removed from the N-terminus or having an  
CC amino acid substitution at position 8 or 706 of the sequence. A  
CC novel method for PCR uses an enzymatically active DNA polymerase  
CC that has at least 80% identity in its amino acid sequence to the Tt  
CC DNA polymerase and has an exonuclease activity removed. Kits for  
CC PCR, strand displacement amplification of DNA, DNA sequencing,  
CC and reverse transcription-PCR are claimed that utilize such Exo-  
CC DNA polymerases. The Tt DNA polymerase is stable at elevated  
CC temperatures, e.g. 70+ deg C. Removal of its exonuclease activity  
CC does not affect its ability to replicate DNA or its thermostability.  
XX

XX SQ Sequence 872 AA;

Query Match 4.8%; Score 91.5; DB 20; Length 872;  
Best Local Similarity 17.5%; Pred. No. 5.7;  
Matches 74; Conservative 70; Mismatches 141; Indels 139; Gaps 19;

QY 14 LKIALKQSLSPFQNSQSLORAYSTPYVYRILOKENKEKQALAHKICISILEFFKN 73  
DB 224 IKENIENNKELAINSKELATIKEDIPEDFEYKVKFNEEK-----LLELPNK 273  
QY 74 LLFVHLLSLKNQREGSGTDMVSTPFFNRNLWYLLSSRFLSKSYCPRFPLDYLEAF 133  
DB 274 LEFSLDNIKE-----SSIEVD-----NHKVKWSK-----VDIKELV 309  
QY 134 GLLSDFLDHQAVIKFELETHFSYYPV-----SGFVAPHOYLSLLQDRYFFIASV 183  
DB 310 TLQD-----NENIAFYELIYEGEIKKIASFGKOTVYIDVFQTE-----DL 351  
QY 184 MRTLKDNFSLTPDLIHDL-----GHVPWLLHPSFSEFFINMGRFLT 226  
DB 352 KEIFEKEDFEFTTHEIKDFLVLRLSYKGIECKSKYIDTAVMAYLLNPSESNY--DLDRVLK 409  
QY 227 KVIEKVOALPS-----KKORIQTLOSNLIA---IVRCFW-----257  
DB 410 KYL-KVD-VPSYEGIFGKGRDKKKEIDENILADYICRVCVYLFDLKELMNFIEMDM 467  
QY 258 -----FTVESGIEHNHGRKAYGAVLSSP-QELGHAFIDNVRVLPLELDQIIR---LPFN 309  
DB 468 KKLLEIEMLPEVLKSMESVSGPTLDEKVELKSKIDDRIGEI---LDKIYKAGYQFN 524  
QY 310 TSTPOE-TLFSIRHFDLVELTSLKSWMLDQGLLESIPLYN-----QEKY 353  
DB 525 VNSPQLSEFLFEKLNLPVTKTKYGTSTDSVQLVPYNDIVSDIIEYRQTLKSTY 584  
QY 354 LSGF 357  
DB 585 IDGF 588

RESULT 49

AAV31815  
ID AAY31815 standard; Protein; 872 AA.

XX AAY31815;

XX 06-DEC-1999 (first entry)

XX Thermoanaerobacter thermohydrosulfuricus DNA polymerase D8A mutant.

XX DNA polymerase; thermostable enzyme; amplification; sequencing;  
KW PCR; reverse transcription; exonuclease; mutant.  
XX  
OS Thermoanaerobacter thermohydrosulfuricus.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 8 /note= "wild-type Asp is substituted by Ala"  
FT  
XX WO9947539-A1.  
PN  
XX 23-SEP-1999.  
PD  
XX 17-MAR-1999; 99WO-US05612.  
PF  
XX 18-MAR-1998; 98US-0044106.  
PR  
XX (AMSH ) AMERSHAM PHARMACIA BIOTECH INC.  
PA  
XX Mamone JA, Davis M, Sha D;  
PI WPI; 1999-562049/47.  
DR  
XX New DNA polymerase from Thermoanaerobacter thermohydrosulfuricus, used  
PT for DNA amplification, sequencing, polymerase chain reaction and  
PT reverse transcription  
XX  
XX Claim 5; Page 7; 71pp; English.  
XX  
CC This sequence represents a claimed mutant DNA polymerase of  
CC Thermoanaerobacter thermohydrosulfuricus (Tt) in which the Asp-8  
CC residue of the native enzyme is replaced by Ala. This amino acid  
CC substitution removes the exonuclease activity of the DNA  
CC polymerase. Claimed DNA polymerases have their exonuclease  
CC activity removed by N-terminal deletion or amino acid substitution.  
CC A novel method for PCR uses an enzymatically active DNA polymerase  
CC that has at least 80% identity in its amino acid sequence to the Tt  
CC DNA polymerase and has an exonuclease activity removed. Kits for  
CC PCR, strand displacement amplification of DNA, DNA sequencing,  
CC and reverse transcription-PCR are claimed that utilize such Exo-  
CC DNA polymerases. The Tt DNA polymerase is stable at elevated  
CC temperatures, e.g. 70+ deg C. Removal of its exonuclease activity  
CC does not affect its ability to replicate DNA or its thermostability.  
CC Note: the present sequence is not shown in the specification but is  
CC derived from the Tt DNA polymerase sequence given in figure 1.  
XX  
XX SQ Sequence 872 AA;

Query Match 4.8%; Score 91.5; DB 20; Length 872;  
Best Local Similarity 17.5%; Pred. No. 5.7;  
Matches 74; Conservative 70; Mismatches 141; Indels 139; Gaps 19;

QY 14 LKIALKQSLSPFQNSQSLORAYSTPYVYRILOKENKEKQALAHKICISILEFFKN 73  
DB 224 IKENIENNKELAINSKELATIKEDIPEDFEYKVKFNEEK-----LLELPNK 273  
QY 74 LLFVHLLSLKNQREGSGTDMVSTPFFNRNLWYLLSSRFLSKSYCPRFPLDYLEAF 133  
DB 274 LEFSLDNIKE-----SSIEVD-----NHKVKWSK-----VDIKELV 309  
QY 134 GLLSDFLDHQAVIKFELETHFSYYPV-----SGFVAPHOYLSLLQDRYFFIASV 183  
DB 310 TLQD-----NENIAFYELIYEGEIKKIASFGKOTVYIDVFQTE-----DL 351  
QY 184 MRTLKDNFSLTPDLIHDL-----GHVPWLLHPSFSEFFINMGRFLT 226  
DB 352 KEIFEKEDFEFTTHEIKDFLVLRLSYKGIECKSKYIDTAVMAYLLNPSESNY--DLDRVLK 409  
QY 227 KVIEKVOALPS-----KKORIQTLOSNLIA---IVRCFW-----257  
DB 410 KYL-KVD-VPSYEGIFGKGRDKKKEIDENILADYICRVCVYLFDLKELMNFIEMDM 467

QY 258 ----FTVESGLIENHGRKAYGAVLISSP-OELGHAFIDNVAVLPLEDOIIR---LPFN 309  
 Db 468 KLLLEIEMPLVEVLSKMEVSGFTLDKEVLELSQKIDDRIGEI---LDKIYKAGYQFN 524  
 QY 310 TSTPOE-TLFSIRHPELVELTSKLEWMLDOGLLESIPLYN-----OEKY 353  
 Db 525 VNSPKQSEFLPEKLNLPVKKTKGTGSTDSEVLEQLVPYNDIVSDIIEYRQLTKLKSTY 584  
 QY 354 LSGP 357  
 Db 585 IDGF 588  
 RESULT 50  
 AAY31816  
 ID AAY31816 standard; Protein; 872 AA.  
 AC AAY31816;  
 XX  
 DT 06-DEC-1999 (first entry)  
 XX  
 DE T. thermohydrosulfuricus DNA polymerase F706Y mutant.  
 XX  
 KM DNA polymerase; thermostable enzyme; amplification; sequencing;  
 XX PCR; reverse transcription; exonuclease; mutant.  
 XX  
 OS Thermoanaerobacter thermohydrosulfuricus.  
 XX Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 706  
 FT /note= "wild-type phe is substituted by Tyr"  
 XX  
 PN WO9947539-A1.  
 XX  
 PD 23-SEP-1999.  
 XX  
 PF 17-MAR-1999; 99WO-US05612.  
 XX  
 PR 18-MAR-1998; 98US-0044106.  
 XX  
 PA (AMSH ) AMERSHAM PHARMACIA BIOTECH INC.  
 XX  
 PI Mamone JA, Davis M, Sha D;  
 XX  
 DR WPI; 1999-562049/47.  
 XX  
 PT New DNA polymerase from Thermoanaerobacter thermohydrosulfuricus, used  
 PT for DNA amplification, sequencing, polymerase chain reaction and  
 PT reverse transcription -  
 XX  
 PS Claim 23; Page -; 71pp; English.  
 XX  
 CC This sequence represents a claimed mutant DNA polymerase of  
 CC Thermoanaerobacter thermohydrosulfuricus (Tt) in which the phe-706  
 CC residue of the native enzyme is replaced by Tyr. This amino acid  
 CC substitution removes the exonuclease activity of the DNA  
 CC polymerase. Claimed DNA polymerases have their exonuclease  
 CC activity removed by N-terminal deletion or amino acid substitution.  
 CC A novel method for PCR uses an enzymatically active DNA polymerase  
 CC that has at least 80% identity in its amino acid sequence to the Tt  
 CC DNA polymerase and has an exonuclease activity removed. Kits for  
 CC PCR, strand displacement amplification of DNA, DNA sequencing,  
 CC and reverse transcription PCR are claimed that utilize such exo-  
 CC DNA polymerases. The Tt DNA polymerase is stable at elevated  
 CC temperatures, e.g. 70+ deg C. Removal of its exonuclease activity  
 CC does not affect its ability to replicate DNA or its thermostability.  
 CC Note: the present sequence is not shown in the specification but is  
 CC derived from the Tt DNA polymerase sequence given in figure 1.  
 XX  
 SQ Sequence 872 AA;

Query Match 4.8%; Score 91.5; DB 20; Length 872;  
 Best Local Similarity 17.5%; Pred. No. 5.7; 141; Indels 139; Gaps 19;  
 Matches 74; Conservative 70; Mismatches 141; Indels 139; Gaps 19;  
 QY 14 LKTLAKRGSLSPFQNSGLORAYSTYSYRIILQENKKEQALARKKISIIIEFFN 73  
 Db 224 IKENIENKRLAIVMSKGLATIGDPIEDIEEYKKEDEK-----LLELFNK 273  
 QY 74 LLPVHLISLKNORBCGCTDMAVSTPPFNMLYRLSSRFSLSKSYCPREFLDYLEAF 133  
 Db 274 LEFSLIDNIKE-----SSIEIVD-----NHKVERKSK-----VDIKELV 309  
 QY 134 GLISDFLDHQAVIKFELETHFSYYPV-----SGFVAPHOYLISLQDRYFPIASV 183  
 Db 310 TLIDP-----NRNIAFPYLIYEGEIKKIAFSFGKDTYIVDFOTE-----DL 351  
 QY 184 MRTLDKDNFSLTPDLIHDL-----GHVPMLHPSFSEFFINMGRLEPT 226  
 Db 352 KEIPEKEDFEFTTHIKDPLVRLSYKGIECKXYIDTAVAVYLLNPESNY--DLDRVLK 409  
 QY 227 KVIKVOALPS-----KKORIOTLQSNLIA--IVRCFM----- 257  
 Db 410 KYL-KVD-VPSYSGIRGKGRDKKKIIEIDENIADYICSCYVIFDLKEXLNFIEKMD 467  
 QY 258 ----FTVESGLIENHGRKAYGAVLISSP-OELGHAFIDNVAVLPLEDOIIR---LPFN 309  
 Db 468 KLLLEIEMPLVEVLSKMEVSGFTLDKEVLELSQKIDDRIGEI---LDKIYKAGYQFN 524  
 QY 310 TSTPOE-TLFSIRHPELVELTSKLEWMLDOGLLESIPLYN-----OEKY 353  
 Db 525 VNSPKQSEFLPEKLNLPVKKTKGTGSTDSEVLEQLVPYNDIVSDIIEYRQLTKLKSTY 584  
 QY 354 LSGP 357  
 Db 585 IDGF 588

Search completed: January 9, 2003, 15:27:08  
 Job time : 77 secs